

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
12 August 2004 (12.08.2004)

PCT

(10) International Publication Number  
**WO 2004/067707 A2**

(51) International Patent Classification<sup>7</sup>: **C12N**  
(21) International Application Number:  
PCT/US2003/041335

70810 (US). **FIORETTI, William, C.** [US/US]; 2225 Lak-  
eridge Road Drive, Grapevine, TX 76051 (US). **CADD,**  
**Gary, G.** [US/US]; 501 Turner Road, Apartment 1111,  
Grapevine, TX 76051 (US).

(22) International Filing Date:  
24 December 2003 (24.12.2003)

(74) Agent: **PRATT, John, S.**; Kilpatrick Stockton LLP, Suite  
2800, 1100 Peachtree Street, Atlanta, GA 30309 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/441,377 21 January 2003 (21.01.2003) US  
60/441,381 21 January 2003 (21.01.2003) US  
60/441,392 21 January 2003 (21.01.2003) US  
60/441,405 21 January 2003 (21.01.2003) US  
60/441,447 21 January 2003 (21.01.2003) US  
60/441,502 21 January 2003 (21.01.2003) US  
10/609,019 26 June 2003 (26.06.2003) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR,  
CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD,  
GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,  
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN,  
MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU,  
SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA,  
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (BW, GH,  
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE,  
SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA,  
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (*for all designated States except US*): **TRANS-  
GENRES, LLC** [US/US]; 1750 Valley View Lane, Suite  
100, Dallas, TX 75243 (US). **THE BOARD OF SUPER-  
VISORS OF LOUISIANA STATE UNIVERSITY AND  
AGRICULTURAL AND MECHANICAL COLLEGE**  
[US/US]; P.O. Box 25055, Baton Rouge, LA 70894 (US).

**Published:**

— *without international search report and to be republished  
upon receipt of that report*

(72) Inventors; and  
(75) Inventors/Applicants (*for US only*): **COOPER, Richard,**  
**K.** [US/US]; 111 Pecan Meadow Drive, Baton Rouge, LA

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

(54) Title: ADMINISTRATION OF TRANSPOSON-BASED VECTORS TO REPRODUCTIVE ORGANS



(57) Abstract: Methods and compositions for the administration of transposon-based vectors to the reproductive organs of animals and the creation of transgenic animals. Preferred methods involve administration of the transposon-based vectors to the lumen of the oviduct of an avian, expression of a vector derived transgene in the avian, and deposition of the resultant polypeptide in an egg. This invention allows for large amounts of protein to be deposited in the egg.

WO 2004/067707 A2

1

## 10

15

## 20

25

## 30

35

Improved gene delivery technologies are also needed for the treatment of disease in animals and humans. Many diseases and conditions can be treated with gene-delivery technologies, which provide a gene of interest to a patient suffering from the disease or the condition. An example of such disease is Type 1 diabetes.

5 Type 1 diabetes is an autoimmune disease that ultimately results in destruction of the insulin producing  $\beta$ -cells in the pancreas. Although patients with Type 1 diabetes may be treated adequately with insulin injections or insulin pumps, these therapies are only partially effective. Insulin replacement, such as via insulin injection or pump administration, cannot fully reverse the defect in the vascular endothelium found in

10 the hyperglycemic state (Pieper et al., 1996. Diabetes Res. Clin. Pract. Suppl. S157-S162). In addition, hyper- and hypoglycemia occurs frequently despite intensive home blood glucose monitoring. Finally, careful dietary constraints are needed to maintain an adequate ratio of calories consumed. This often causes major psychosocial stress for many diabetic patients. Development of gene therapies

15 providing delivery of the insulin gene into the pancreas of diabetic patients could overcome many of these problems and result in improved life expectancy and quality of life.

Several of the prior art gene delivery technologies employed viruses that are associated with potentially undesirable side effects and safety concerns. The majority

20 of current gene-delivery technologies useful for gene therapy rely on virus-based delivery vectors, such as adeno and adeno-associated viruses, retroviruses, and other viruses, which have been attenuated to no longer replicate. (Kay, M.A., et al. 2001. Nature Medicine 7:33-40).

There are multiple problems associated with the use of viral vectors. Firstly,

25 they are not tissue-specific. In fact, a gene therapy trial using adenovirus was recently halted because the vector was present in the patient's sperm (Gene trial to proceed despite fears that therapy could change child's genetic makeup. The New York Times, December 23, 2001). Secondly, viral vectors are likely to be transiently incorporated, which necessitates re-treating a patient at specified time intervals. (Kay, M.A., et al. 2001. Nature Medicine 7:33-40). Thirdly, there is a concern that a viral-

30 based vector could revert to its virulent form and cause disease. Fourthly, viral-based vectors require a dividing cell for stable integration. Fifthly, viral-based vectors indiscriminately integrate into various cells, which can result in undesirable germline integration. Sixthly, the required high titers needed to achieve the desired effect have

35 resulted in the death of one patient and they are believed to be responsible for induction of cancer in a separate study. (Science, News of the Week, October 4, 2002).

Accordingly, what is needed is a new method to produce transgenic animals and humans with stably incorporated genes, in which the vector containing those genes does not cause disease or other unwanted side effects. There is also a need for DNA constructs that would be stably incorporated into the tissues and cells of animals and humans, including cells in the resting state that are not replicating. There is a further recognized need in the art for DNA constructs capable of delivering genes to specific tissues and cells of animals and humans.

When incorporating a gene of interest into an animal for the production of a desired protein or when incorporating a gene of interest in an animal or human for the treatment of a disease, it is often desirable to selectively activate incorporated genes using inducible promoters. These inducible promoters are regulated by substances either produced or recognized by the transcription control elements within the cell in which the gene is incorporated. In many instances, control of gene expression is desired in transgenic animals or humans so that incorporated genes are selectively activated at desired times and/or under the influence of specific substances. Accordingly, what is needed is a means to selectively activate genes introduced into the genome of cells of a transgenic animal or human. This can be taken a step further to cause incorporation to be tissue-specific, which prevents widespread gene incorporation throughout a patient's body (animal or human). This decreases the amount of DNA needed for a treatment, decreases the chance of incorporation in gametes, and targets gene delivery, incorporation, and expression to the desired tissue where the gene is needed to function. What is also needed is a rapid expression method for rapidly producing a protein or peptide of interest in eggs and milk of transgenic animals.

25

#### SUMMARY OF THE INVENTION

The present invention addresses the problems described above by providing new, effective and efficient compositions for producing transgenic animals and for treating disease in animals or humans. Transgenic animals include all egg-laying animals and milk-producing animals. Transgenic animals further include but are not limited to avians, fish, amphibians, reptiles, insects, mammals and humans. In another preferred embodiment, the animal is a milk-producing animal, including but not limited to bovine, porcine, ovine and equine animals. In a preferred embodiment, the animal is an avian animal. In another preferred embodiment, the animal is a mammal. Animals are made transgenic through administration of a composition comprising a transposon-based vector designed for incorporation of a gene of interest for production of a desired protein, together with an acceptable carrier. The

35



compositions of the present invention are introduced into the reproductive system of an animal. The compositions of the present invention are administered to a reproductive organ including, but not limited to, an oviduct, an ovary, or into the duct system of the mammary gland. The compositions of the present invention may be administered to a reproductive organ of an animal through the cloaca. The compositions of the present invention may be directly administered to a reproductive organ or can be administered to an artery leading to the reproductive organ. In a preferred embodiment, the compositions of the present invention are introduced into the the reproductive system of an avian animal. In another preferred embodiment, the compositions of the present invention are introduced into the the intramammary duct system of a mammal. A transfection reagent is optionally added to the composition before administration.

The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes the following characteristics: a) one or more modified Kozak sequences at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

The present invention also provides for tissue-specific incorporation and/or expression of a gene of interest. Tissue-specific incorporation of a gene of interest may be achieved by placing the transposase gene under the control of a tissue-specific promoter, whereas tissue-specific expression of a gene of interest may be achieved by placing the gene of interest under the control of a tissue-specific promoter. In some embodiments, the gene of interest is transcribed under the influence of an ovalbumin, or other oviduct specific, promoter. Linking the gene of interest to an oviduct specific

promoter in an egg-laying animal results in synthesis of a desired molecule and deposition of the desired molecule in a developing egg.

The present invention advantageously produces a high number of transgenic animals having a gene of interest stably incorporated. In some embodiments wherein  
5 the transposon-based vector is administered to the ovary, these transgenic animals successfully pass the desired gene to their progeny. Accordingly, the present invention can be used to obtain transgenic animals having the gene of interest incorporated into the germline through transfection of the ovary or the present invention can be used to obtain transgenic animals having the gene of interest  
10 incorporated into the oviduct in a tissue-specific manner. Both types of transgenic animals of the present invention produce large amounts of a desired molecule encoded by the transgene. Transgenic egg-laying animals, particularly avians, produce large amounts of a desired protein that is deposited in the egg for rapid harvest and purification.

15 Any desired gene may be incorporated into the novel transposon-based vectors of the present invention in order to synthesize a desired molecule in the transgenic animals. Proteins, peptides and nucleic acids are preferred desired molecules to be produced by the transgenic animals of the present invention. Particularly preferred proteins are antibody proteins and other immunopharmaceutical proteins.

20 This invention provides a composition useful for the production of transgenic hens capable of producing substantially high amounts of a desired protein or peptide. Entire flocks of transgenic birds may be developed very quickly in order to produce industrial amounts of desired molecules. The present invention solves the problems inherent in the inadequate capacity of fermentation facilities used for bacterial  
25 production of molecules and provides a more efficient and economical way to produce desired molecules. Accordingly, the present invention provides a means to produce large amounts of therapeutic, diagnostic and reagent molecules.

Transgenic chickens are excellent in terms of convenience and efficiency of manufacturing molecules such as proteins and peptides. Starting with a single  
30 transgenic rooster, thousands of transgenic offspring can be produced within a year. (In principle, up to forty million offspring could be produced in just three generations). Each transgenic female is expected to lay at least 250 eggs/year, each potentially containing hundreds of milligrams of the selected protein. Flocks of chickens numbering in the hundreds of thousands are readily handled through

established commercial systems. The technologies for obtaining eggs and fractionating them are also well known and widely accepted. Thus, for each therapeutic, diagnostic, or other protein of interest, large amounts of a substantially pure material can be produced at relatively low incremental cost.

5           A wide range of recombinant peptides and proteins can be produced in transgenic egg-laying animals. Enzymes, hormones, antibodies, growth factors, serum proteins, commodity proteins, biological response modifiers, peptides and designed proteins may all be made through practice of the present invention. For example, rough estimates suggest that it is possible to produce in bulk growth  
10   hormone, insulin, or Factor VIII, and deposit them in egg whites, for an incremental cost in the order of one dollar per gram. At such prices it is feasible to consider administering such medical agents by inhalation or even orally, instead of through injection. Even if bioavailability rates through these avenues were low, the cost of a much higher effective-dose would not be prohibitive.

15           In one embodiment, the egg-laying transgenic animal is an avian. The method of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the  
20   bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

          Accordingly, it is an object of the present invention to provide novel  
25   transposon-based vectors.

          It is another object of the present invention to provide novel transposon-based vectors that encode for the production of desired proteins or peptides in cells.

          It is an object of the present invention to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector.

30           Another object of the present invention is to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector, wherein the transgenic animals produce desired proteins or peptides.

It is further an object of the present invention to provide a method to produce transgenic animals through intraovarian administration of a transposon-based vector that are capable of producing transgenic progeny.

Yet another object of the present invention is to provide a method to produce  
5 transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing a desired molecule, such as a protein, peptide or nucleic acid.

Another object of the present invention is to provide a method to produce transgenic animals through intraoviduct or intraovarian administration of a  
10 transposon-based vector, wherein such administration results in modulation of endogenous gene expression.

It is yet another object of the present invention to provide a method to produce transgenic avians through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing proteins, peptides or nucleic acids.

15 It is another object of the present invention to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector encoding an antibody or a fragment thereof.

Still another object of the present invention is to provide a method to produce transgenic avians through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing proteins or peptides and depositing these  
20 proteins or peptides in the egg.

Another object of the present invention is to provide transgenic avians that contain a stably incorporated transgene.

Still another object of the present invention is to provide eggs containing  
25 desired proteins or peptides encoded by a transgene incorporated into the transgenic avian that produces the egg.

It is further an object of the present invention to provide a method to produce transgenic milk-producing animals through administration of a transposon-based vector that are capable of producing proteins, peptides or nucleic acids.

30 Still another object of the present invention is to provide a method to produce transgenic milk-producing animals through administration of a transposon-based vector that are capable of producing proteins or peptides and depositing these proteins or peptides in their milk.

Another object of the present invention is to provide transgenic milk-producing animals that contain a stably incorporated transgene.

Another object of the present invention is to provide transgenic milk-producing animals that are capable of producing proteins or peptides and depositing  
5 these proteins or peptides in their milk.

Yet another object of the present invention is to provide milk containing desired molecules encoded by a transgene incorporated into the transgenic milk-producing animals that produce the milk.

Still another object of the present invention is to provide milk containing  
10 desired proteins or peptides encoded by a transgene incorporated into the transgenic milk-producing animals that produce the milk.

An advantage of the present invention is that transgenic animals are produced with higher efficiencies than observed in the prior art.

Another advantage of the present invention is that these transgenic animals  
15 possess high copy numbers of the transgene.

Another advantage of the present invention is that the transgenic animals produce large amounts of desired molecules encoded by the transgene.

Still another advantage of the present invention is that desired molecules are produced by the transgenic animals much more efficiently and economically than  
20 prior art methods, thereby providing a means for large scale production of desired molecules, particularly proteins and peptides.

Yet another advantage of the present invention is that the desired proteins and peptides are produced rapidly after making animals transgenic through introduction of the vectors of the present invention.

25 These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and claims.

#### BRIEF DESCRIPTION OF THE FIGURES

30 Figure 1 depicts schematically a transposon-based vector containing a transposase operably linked to a first promoter and a gene of interest operably-linked to a second promoter, wherein the gene of interest and its operably-linked promoter are flanked by insertion sequences (IS) recognized by the transposase. "Pro"



designates a promoter. In this and subsequent figures, the size of the actual nucleotide sequence is not necessarily proportionate to the box representing that sequence.

Figure 2 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ov pro is the ovalbumin promoter, Ov protein is the ovalbumin protein and PolyA is a polyadenylation sequence. The TAG sequence includes a spacer sequence, the gp41 hairpin loop from HIV I and a protease cleavage site.

Figure 3 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ovo pro is the ovomucoid promoter and Ovo SS is the ovomucoid signal sequence. The TAG sequence includes a spacer, the gp41 hairpin loop from HIV I and a protease cleavage site.

Figure 4 depicts schematically a transposon based-vector for expression of an RNAi molecule. "Tet<sub>i</sub> pro" indicates a tetracycline inducible promoter whereas "pro" indicates the pro portion of a prepro sequence as described herein. "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotrans" indicates approximately 60 base pairs of an ovotransferrin gene and "Ovomucin" indicates approximately 60 base pairs of an ovomucin gene.

Figure 5 is a picture of an SDS-PAGE gel wherein a pooled fraction of an isolated proinsulin fusion protein was run in lanes 4 and 6. Lanes 1 and 10 of the gel contain molecular weight standards, lanes 2 and 8 contain non-transgenic chicken egg white, and lanes 3, 5, 7 and 9 are blank.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a new, effective and efficient method of producing transgenic animals, particularly egg-laying animals and milk-producing animals, through administration of a composition comprising a transposon-based vector designed for incorporation of a gene of interest and production of a desired molecule. The transposon-based vectors are administered to a reproductive organ including, but not limited to, an oviduct, an ovary, or into the duct system of the mammary gland. The vectors may be directly administered to a reproductive organ or can be administered to an artery leading to the reproductive organ or to a lymph system proximate to the cells to be genetically altered. The vectors may be administered to a reproductive organ of an animal through the cloaca. One method of direct administration is by injection, and in one embodiment, the lumen of the

magnum of the oviduct is injected with a transposon-based vector. Another method of direct administration is by injection, and in one embodiment, the lumen of the infundibulum of the oviduct is injected with a transposon-based vector. A preferred intrarterial administration is an administration into an artery that supplies the oviduct or the ovary. In some embodiments, administration of the transposon-based vector to an oviduct or an artery that leads to the oviduct results in incorporation of the vector into the epithelial and/or secretory cells of the oviduct. In other embodiments, administration of the transposon-based vector to an ovary or an artery that leads to the ovary or a lymphatic system proximal to the ovary results in incorporation of the vector into an oocyte or a germinal disk inside the ovary.

#### Definitions

It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

The term "antibody" is used interchangeably with the term "immunoglobulin" and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an "antigen" or an "immunogen". The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an "antigenic determinant" or an "epitope". Antigens can be naturally occurring or artificially engineered. Artificially engineered antigens include, but are not limited to, small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes,

Passeriformes, Coraciformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a group of flightless, mostly large, running birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes",  
5 as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any member of the avian family Psittacidae (the single family of the Psittaciformes), distinguished by the short, stout, strongly hooked beak. Avians include all poultry birds, especially chickens, geese, turkeys, ducks and quail. The term "chicken" as used herein denotes  
10 chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat production ("dual-purpose" chickens). The term "chicken" also denotes chickens produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg,  
15 meat, or table egg and meat consumption.

The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcarous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the  
20 production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcarous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

The term "milk-producing animal" refers herein to mammals including, but  
25 not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humans.

The term "gene" is defined herein to include a coding region for a protein, peptide or polypeptide.

30 The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more cells that contain a portion of the transposon-based vector DNA for any

period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the reproductive life of the animal, and most preferably the life of the animal. In a further preferred  
5 embodiment, the animal is an avian.

The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or  
10 organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are  
15 associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

The term "regulatory sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences,  
20 matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

#### Transposon-Based Vectors

While not wanting to be bound by the following statement, it is believed that  
25 the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used  
30 often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration frequencies. Two preferred transposon-based vectors of the present invention in

which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:2) and pTnMod (SEQ ID NO:3).

The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive roosters. Actual frequencies of integration were estimated by either or both comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:1) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:4), AAGATGT (SEQ ID NO:5), ACGATGA (SEQ ID NO:6), AAGATGG (SEQ ID NO:7), GACATGA (SEQ ID NO:8), ACCATGA (SEQ ID NO:9), and ACCATGA (SEQ ID NO:10), ACCATGT (SEQ ID NO:52).

Figure 1 shows a schematic representation of several components of the transposon-based vector. The present invention further includes vectors containing more than one gene of interest, wherein a second or subsequent gene of interest is operably-linked to the second promoter or to a different promoter. It is also to be



understood that the transposon-based vectors shown in the Figures are representative of the present invention and that the order of the vector elements may be different than that shown in the Figures, that the elements may be present in various orientations, and that the vectors may contain additional elements not shown in the  
5 Figures.

#### Transposases and Insertion Sequences

In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase  
10 located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 ( $\gamma\delta$ ), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted  
15 transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:1) at the 3' end of the promoter operably-linked to the transposase; b) a change of the codons for the first  
20 several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

25 Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene increase transcription of the transposase gene, in part, by increasing strand dissociation. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the  
30 transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred

insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

5 In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

10 Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the *Drosophila* P element derived vectors disclosed in U.S. Patent No. 6,291,243; the *Drosophila* mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 15 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), Mobile DNA, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

20 Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type 25 transposons that typically contain insertion sequences of well over 100 base pairs.

While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a rolling replication type transposon allows multiple copies of the transposon/transgene 30 to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use

of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

5           Stop Codons and PolyA Sequences

In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest.

10           As used herein an "effective polyA sequence" refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a  
15           gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that  
20           precede the A polynucleotide string and thereby separate the stop codon from the A polynucleotide string. In one embodiment of the present invention, the polyA sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:11 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence  
25           shown in SEQ ID NO:12. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:13. A chicken optimized polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

It is a surprising discovery of the present invention that such an avian  
30           optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to

an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 60 base pairs that precede the A polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

#### Promoters and Enhancers

The first promoter operably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters,  $\beta$ -actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID NO:14. The G6P promoter sequence may be deduced from a rat G6P gene untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothioneine promoter.

Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. Genes

and Development, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. Mol. Ther. 1:159-164); estrogen-based inducible systems (Brasemann, S. et al. 1993. Proc. Natl. Acad. Sci. 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4  
5 binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. Proc. Natl. Acad. Sci. 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces  
10 dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. J. Chem. Biol. 3:731-738; Fan, L. et al. 1999. Hum. Gene Ther. 10:2273-2285; Shariat, S.F. et al. 2001. Cancer Res. 61:2562-2571; Spencer, D.M. 1996. Curr. Biol. 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known  
15 to those of skill in the art.

Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. Circ Res. 91(12):1151-9); ubiquitin C promoter (Biochim Biophys Acta, 2003. Jan. 3;1625(1):52-63); Hsf2  
20 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. Mol. Cell Biol. 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, J. Urol. 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. Proc. Natl. Acad. Sci. 99(24):15456-61); neural and liver ceramidase  
25 gene promoters (Okino N. et al., 2002. Biochem. Biophys. Res. Commun. 299(1):160-6); PSP94 gene promoter/enhancer (Gabril M.Y. et al., 2002. Gene Ther. 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. Biol. Pharm. Bull. 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. Blood October 24, 2002); and, IL-10 promoter (Brenner S., et al., 2002. J. Biol. Chem.  
30 December 18, 2002).

Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding



protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very low-density lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavin-binding protein, biotin-binding protein (Awade, 1996. Z. Lebensm. Unters. Forsch. 5 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific 10 promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin; ovotransferrin (conalbumin); ovomucoid; 01, 02, 03, 04 or 05 avidin; ovomucin; g2 ovoglobulin; g3 ovoglobulin; ovoflavoprotein; and ovostatin (ovomacroglobin) promoters.

When germline transformation occurs via intraovarian administration, liver- 15 specific promoters may be operably-linked to the gene of interest to achieve liver-specific expression of the transgene. Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like 20 growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, 25 but not limited to,  $\beta$  lactoglobulin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

When germline transformation occurs via intraovarian administration, immune system-specific promoters may be operably-linked to the gene of interest to achieve immune system-specific expression of the transgene. Accordingly, promoters 30 associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type

2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved  
5 from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural  
10 backbone, are all encompassed by the current invention.

Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a  
15 chicken ovalbumin promoter (SEQ ID NO:15). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:1).

As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-  
20 linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer and a promoter modified to eliminate repressive regulatory effects are referred to herein as an "enhanced promoter." The enhancers contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer,  
25 but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a chicken ovalbumin enhancer from GenBank  
30 accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:16.

Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of

an enhanced promoter or promoter. An exemplary 5' untranslated region is provided in SEQ ID NO:17. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:18.

5 In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is  
10 incorporated into the germline generally, the gene of interest may only be expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary, to an artery leading to the ovary or to a lymphatic system or fluid proximal to the ovary.

15 It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

20 When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the ovalbumin promoter provides for activation of the transposase gene and incorporation  
25 of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is  
30 desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest, to an artery leading to the tissue of interest or to fluids surrounding the tissue of interest. In a preferred embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the oviduct or an artery leading to the oviduct. In a further preferred embodiment, administration is achieved by

direct injection into the lumen of the magnum or the infundibulum of the oviduct. Indirect administration to the oviduct may occur through the cloaca.

Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobulin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

#### Directing Sequences

In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences. An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, that directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene

therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. One example of such a transposon-based vector is provided in Figure 3 wherein the gene of interest is operably linked to the ovomucoid signal sequence. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is shown in Figure 2 wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:19. In another embodiment, the signal sequence is a modified ovalbumin signal sequence including a sequence shown in SEQ ID NO:20 or SEQ ID NO:21.

As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:22. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

#### Genes of Interest Encoding Desired Proteins

A gene of interest selected for stable incorporation is designed to encode any desired protein or peptide or to regulate any cellular response. In some embodiments,



the desired proteins or peptides are deposited in an egg or in milk. It is to be understood that the present invention encompasses transposon-based vectors containing multiple genes of interest. The multiple genes of interest may each be operably-linked to a separate promoter and other regulatory sequence(s) or may all be  
5 operably-linked to the same promoter and other regulatory sequences(s). In one embodiment, multiple gene of interest are linked to a single promoter and other regulatory sequence(s) and each gene of interest is separated by a cleavage site or a pro portion of a signal sequence. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the  
10 third base of each codon is changed to an A or a T without changing the corresponding amino acid.

Protein and peptide hormones are a preferred class of proteins in the present invention. Such protein and peptide hormones are synthesized throughout the endocrine system and include, but are not limited to, hypothalamic hormones and  
15 hypophysiotropic hormones, anterior, intermediate and posterior pituitary hormones, pancreatic islet hormones, hormones made in the gastrointestinal system, renal hormones, thymic hormones, parathyroid hormones, adrenal cortical and medullary hormones. Specifically, hormones that can be produced using the present invention include, but are not limited to, chorionic gonadotropin, corticotropin, erythropoietin,  
20 glucagons, IGF-1, oxytocin, platelet-derived growth factor, calcitonin, follicle-stimulating hormone, luteinizing hormone, thyroid-stimulating hormone, insulin, gonadotropin-releasing hormone and its analogs, vasopressin, octreotide, somatostatin, prolactin, adrenocorticotrophic hormone, antidiuretic hormone, thyrotropin-releasing hormone (TRH), growth hormone-releasing hormone (GHRH),  
25 dopamine, melatonin, thyroxin (T<sub>4</sub>), parathyroid hormone (PTH), glucocorticoids such as cortisol, mineralocorticoids such as aldosterone, androgens such as testosterone, adrenaline (epinephrine), noradrenaline (norepinephrine), estrogens such as estradiol, progesterone, glucagons, calcitrol, calciferol, atrial-natriuretic peptide, gastrin, secretin, cholecystokinin (CCK), neuropeptide Y, ghrelin, PYY<sub>3-36</sub>,  
30 angiotensinogen, thrombopoietin, and leptin. By using appropriate polynucleotide sequences, species-specific hormones may be made by transgenic animals.

In one embodiment of the present invention, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become

linked together to form the insulin molecule. Figures 2 and 3 are schematics of transposon-based vector constructs containing a proinsulin gene operably-linked to an ovalbumin promoter and ovalbumin protein or an ovomucoid promoter and ovomucoid signal sequence, respectively. In these embodiments, proinsulin is  
5 expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:23, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

Serum proteins including lipoproteins such as high density lipoprotein (HDL), HDL-Milano and low density lipoprotein, albumin, clotting cascade factors, factor  
10 VIII, factor IX, fibrinogen, and globulins are also included in the group of desired proteins of the present invention. Immunoglobulins are one class of desired globulin molecules and include but are not limited to IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains and fragments thereof; Fc fragments, and Fab fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies,  
15 human antibodies, humanized antibodies, and hybrid antibodies. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies  
20 against proteins associated with cancer-related molecules, such as anti-her 2, or anti-CA125. Accordingly, the present invention encompasses a transposon-based vector containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain. Further, more than one gene encoding for more than one antibody may be administered in one or more transposon-based vectors of the present invention. In  
25 this manner, an egg may contain more than one type of antibody in the egg white, the egg yolk or both. In one embodiment, a transposon-based vector contains a heavy Ig chain and a light Ig chain, both operably linked to a promoter.

Antibodies used as therapeutic reagents include but are not limited to antibodies for use in cancer immunotherapy against specific antigens, or for providing  
30 passive immunity to an animal or a human against an infectious disease or a toxic agent. Antibodies used as diagnostic reagents include, but are not limited to antibodies that may be labeled and detected with a detector, for example antibodies with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a

specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken IgG. Such labeled antibodies are known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some  
5 of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety.

Antibodies produced with using the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western  
10 blots, dot blots, ELISA, immunoaffinity columns and other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

Antibodies that may be made with the practice of the present invention  
15 include, but are not limited to primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, anti-hormone antibodies, anti-hypophysiotropic peptides, antibodies against non-natural antigens, anti-anterior pituitary hormone antibodies, anti-posterior pituitary hormone antibodies, anti-venom antibodies, anti-tumor marker antibodies,  
20 antibodies directed against epitopes associated with infectious disease, including, anti-viral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotypic, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme  
25 activity.

Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine  $\gamma$ -Globulin, Serum; Bovine IgG, Plasma; Chicken  $\gamma$ -Globulin, Serum; Human  $\gamma$ -Globulin, Serum; Human IgA, Plasma; Human IgA<sub>1</sub>, Myeloma; Human IgA<sub>2</sub>, Myeloma; Human IgA<sub>2</sub>,  
30 Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, F(ab')<sub>2</sub> Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG<sub>1</sub>, Myeloma; Human IgG<sub>2</sub>, Myeloma; Human IgG<sub>3</sub>, Myeloma; Human IgG<sub>4</sub>, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain  $\kappa$ , Urine; Human Immunoglobulin, Light

Chains  $\kappa$  and  $\lambda$ , Plasma; Mouse  $\gamma$ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit  $\gamma$ -Globulin, Serum; Rabbit IgG, Plasma; and Rat  $\gamma$ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows  
 5 specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

A further non-limiting list of antibodies that recognize other antibodies is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat  $\gamma$ -Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig  
 10  $\gamma$ -Globulin (Goat); Anti-Human Ig, Light Chain, Type  $\kappa$  Specific; Anti-Human Ig, Light Chain, Type  $\lambda$  Specific; Anti-Human IgA,  $\alpha$ -Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE,  $\epsilon$ -Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human  
 15 IgG,  $\gamma$ -Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG<sub>1</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fd Fragment Specific; Anti-Human IgG<sub>3</sub>, Hinge Specific; Anti-Human IgG<sub>4</sub>, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human  
 20 IgM,  $\mu$ -Chain Specific; Anti-Mouse IgE,  $\epsilon$ -Chain Specific; Anti-Mouse  $\gamma$ -Globulin (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat) F(ab')<sub>2</sub> Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM,  $\mu$ -Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit  $\gamma$ -Globulin (Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit  
 25 IgG, H & L Chain Specific (Goat); Anti-Rat  $\gamma$ -Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey  $\gamma$ -Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix  
 30 Pharmaceuticals, Inc. ([www.phoenixpeptide.com](http://www.phoenixpeptide.com); 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA (St. Louis, MO [www.sigma-aldrich.com](http://www.sigma-aldrich.com)), Cappel ICN (Irvine, California, [www.icnbiomed.com](http://www.icnbiomed.com)), and Calbiochem (La Jolla, California, [www.calbiochem.com](http://www.calbiochem.com)), which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies

may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired antibody. Antibodies made by the transgenic animals of the present invention include antibodies that may be used as therapeutic reagents, for example in cancer immunotherapy against specific antigens, as diagnostic reagents and as laboratory reagents for numerous applications including immunoneutralization, radioimmunoassay, western blots, dot blots, ELISA, immunoprecipitation and immunoaffinity columns. Some of these antibodies include, but are not limited to, antibodies which bind the following ligands: adrenomedulin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiotensin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokineticin, thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:24 amino acid sequence, SEQ ID NO:25, nucleotide sequence), estrogen, testosterone, corticosteroids, mineralocorticoids, thyroid hormone, thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, and prohormones, propeptides, splice variants, fragments and analogs thereof.

The following is yet another non-limiting list of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody,



anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, anti-hepatitis B antibody, anti-HER-2 antibody, anti-idiotypic antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, 5 anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNF $\alpha$ , antibody directed against GPIIb/IIIa receptor on human platelets, gamma 10 globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against *S aureus*, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the  $\alpha$  subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R  $\alpha$ -chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized 15 mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX- 20 33 antibody to Fc $\gamma$ R1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3) antibody, OKT3 ("muromomab-CD3") antibody, PRO 542 antibody, ReoPro 25 ("abciximab") antibody, and TNF-IgG fusion protein.

The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary 30 skill in the art. Specialty antibodies designed for binding specific antigens may also be made in transgenic animals using the transposon-based vectors of the present invention.

Production of a monoclonal antibody using the transposon-based vectors of the present invention can be accomplished in a variety of ways. In one embodiment, two vectors may be constructed: one that encodes the light chain, and a second vector that encodes the heavy chain of the monoclonal antibody. These vectors may then be  
5 incorporated into the genome of the target animal by methods disclosed herein. In an alternative embodiment, the sequences encoding light and heavy chains of a monoclonal antibody may be included on a single DNA construct. For example, the coding sequence of light and heavy chains of a murine monoclonal antibody that show specificity for human seminoprotein can be expressed using transposon-based  
10 constructs of the present invention (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

Further included in the present invention are proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The  
15 immune system proteins and peptides proteins that can be made in transgenic animals using the transposon-based vectors of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interleukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF- $\alpha$ , and TNF- $\beta$ . Other cytokines  
20 included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5.

Lytic peptides such as p146 are also included in the desired molecules of the  
25 present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:24. The present invention also encompasses a transposon-based vector comprising a p146 nucleic acid comprising a polynucleotide sequence of SEQ ID NO:25.

Enzymes are another class of proteins that may be made through the use of the  
30 transposon-based vectors of the present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI, hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase

and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in  
5 activation of the enzyme.

Extracellular matrix proteins are one class of desired proteins that may be made through the use of the present invention. Examples include but are not limited to collagen, fibrin, elastin, laminin, and fibronectin and subtypes thereof. Intracellular proteins and structural proteins are other classes of desired proteins in the present  
10 invention.

Growth factors are another desired class of proteins that may be made through the use of the present invention and include, but are not limited to, transforming growth factor- $\alpha$  ("TGF- $\alpha$ "), transforming growth factor- $\beta$  (TGF- $\beta$ ), platelet-derived growth factors (PDGF), fibroblast growth factors (FGF), including FGF acidic isoforms 1 and 2, FGF basic form 2 and FGF 4, 8, 9 and 10, nerve growth factors  
15 (NGF) including NGF 2.5s, NGF 7.0s and beta NGF and neurotrophins, brain derived neurotrophic factor, cartilage derived factor, growth factors for stimulation of the production of red blood cells, growth factors for stimulation of the production of white blood cells, bone growth factors (BGF), basic fibroblast growth factor, vascular  
20 endothelial growth factor (VEGF), granulocyte colony stimulating factor (G-CSF), insulin like growth factor (IGF) I and II, hepatocyte growth factor, glial neurotrophic growth factor (GDNF), stem cell factor (SCF), keratinocyte growth factor (KGF), transforming growth factors (TGF), including TGFs alpha, beta, beta1, beta2, beta3, skeletal growth factor, bone matrix derived growth factors, bone derived growth  
25 factors, erythropoietin (EPO) and mixtures thereof.

Another desired class of proteins that may be made may be made through the use of the present invention include, but are not limited to, leptin, leukemia inhibitory factor (LIF), tumor necrosis factor alpha and beta, ENBREL, angiostatin, endostatin, thrombospondin, osteogenic protein-1, bone morphogenetic proteins 2 and 7,  
30 osteonectin, somatomedin-like peptide, and osteocalcin.

Yet another desired class of proteins are blood proteins or clotting cascade protein including albumin, Prekallikrein, High molecular weight kininogen (HMWK) (contact activation cofactor; Fitzgerald, Flaubeac Williams factor), Factor I

(Fibrinogen), Factor II (prothrombin), Factor III (Tissue Factor), Factor IV (calcium), Factor V (proaccelerin, labile factor, accelerator (Ac-) globulin), Factor VI (Va) (accelerin), Factor VII (proconvertin), serum prothrombin conversion accelerator (SPCA), cothromboplastin), Factor VIII (antihemophilic factor A, antihemophilic globulin (AHG)), Factor IX (Christmas Factor, antihemophilic factor B, plasma thromboplastin component (PTC)), Factor X (Stuart-Prower Factor), Factor XI (Plasma thromboplastin antecedent (PTA)), Factor XII (Hageman Factor), Factor XIII (rotransglutaminase, fibrin stabilizing factor (FSF), fibrinolygase), von Willebrand factor, Protein C, Protein S, Thrombomodulin, Antithrombin III.

10 A non-limiting list of the peptides and proteins that may be made may be made through the use of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. ([www.phoenixpeptide.com](http://www.phoenixpeptide.com); 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St.Louis, MO [www.sigma-aldrich.com](http://www.sigma-aldrich.com)), Cappel ICN (Irvine, California, [www.icnbiomed.com](http://www.icnbiomed.com)), and Calbiochem (La Jolla, California, [www.calbiochem.com](http://www.calbiochem.com)).  
15 The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons  
20 that encode for each amino acid in the desired protein or peptide.

Some of these desired proteins or peptides that may be made through the use of the present invention include but are not limited to the following: adrenomedulin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin,  
25 somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM,  
30 PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiostatin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide,

peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokineticin, thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:24, amino acid sequence, SEQ ID NO:25, nucleotide sequence), thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, apolipoproteins, apolipoprotein A, apolipoprotein B, and prohormones, propeptides, splice variants, fragments and analogs thereof.

Other desired proteins that may be made by the transgenic animals of the present invention include bacitracin, polymixin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, anti-hepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin,  $\alpha$ -galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin), goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus



(MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsonin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNF $\alpha$ , mAb  
5 against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectapase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, fab fragment of anti-VEGF  
10 Ab, APO2 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF- $\beta$  (transforming growth factor- $\beta$ ), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase,  $\alpha$ -L-iduronidase,  $\alpha$ -L-galactosidase (humanglutamic acid decarboxylase, acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV  
15 vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotypic antibody to GD3 glycolipid, antibacterial peptide against *H. pylori*, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized  
20 recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on *Escherichia coli* strains), proteins of lyme disease vaccine against *B. burgdorferi* protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), *Streptococcus pneumoniae* vaccine, Medi-507 mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb  
25 to Fc $\gamma$ R1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bi-specific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R  $\alpha$ -chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A),  
30 humanized mAb against the epidermal growth receptor-2, humanized mAb against the  $\alpha$  subunit of the interleukin-2 receptor, low molecular weight heparin, anti-hemophilic factor, and bactericidal/permeability-increasing protein (r-BPI).

The peptides and proteins made using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of these

labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected protein or peptide. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label protein.

It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

The present invention may also be used to produce desired molecules other than proteins and peptides including, but not limited to, lipoproteins such as high density lipoprotein (HDL), HDL-Milano, and low density lipoprotein, lipids, carbohydrates, siRNA and ribozymes. In these embodiments, a gene of interest encodes a nucleic acid molecule or a protein that directs production of the desired molecule.

The present invention further encompasses the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins. In a preferred embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin. In one embodiment, a transposon-based vector containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi). In other embodiments, a transposon-based vector encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. One exemplary construct is provided in Figure 4 wherein "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotrans" indicates approximately 60 base pairs of an ovotransferrin gene and "Ovomucin" indicates approximately 60 base pairs of an ovomucin gene. These ovalbumin, ovotransferrin and ovomucin can be from any avian species, and in some embodiments, are from a chicken or quail. The term "pro" indicates the pro portion of a prepro sequence. One exemplary prepro sequence is that of cecropin and comprising base pairs 563-733 of the Cecropin cap site and Prepro provided in

Genbank accession number X07404. Additional cecropin prepro and pro sequences are provided in SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, and SEQ ID NO:51. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production.

- 5 Endogenous egg white production can be inhibited in an avian at any time, but is preferably inhibited preceding, or immediately preceding, the harvest of eggs.

#### Modified Desired Proteins and Peptides

- "Proteins", "peptides," "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds  
10 formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated  
15 N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the  
20 carboxyl group of an amino acid at any other location within the protein.

Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

- 25 The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide  
30 bond mimetic includes peptide backbone modifications well known to those skilled in the art.

Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less

than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids

5 that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 10 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid  
15 being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

20 Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which  
25 can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the  
30

carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

5 Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

10 Group I includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.

Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.

15 Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro,  $\text{—NH}_2$ , methoxy, ethoxy and  $\text{—CN}$ . Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

20 Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine,  $\text{—CO—NH—}$  alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the  
25 side chain  $\text{—(CH}_2)_3\text{—COOH}$ , an ester thereof (substituted or unsubstituted aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and  
30 asparagine.

Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine,



arginine and ornithine. A homolog of an amino acid includes from 1 to about 3 additional or subtracted methylene units in the side chain.

Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with —OH or —SH, for example, —CH<sub>2</sub>CH<sub>2</sub>OH, —CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>OH or —CH<sub>2</sub>CH<sub>2</sub>OHCH<sub>3</sub>. Preferably, Group VI includes serine, cysteine or threonine.

In another aspect, suitable substitutions for amino acid residues include “severe” substitutions. A “severe substitution” is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for glycine, a D amino acid for the corresponding L amino acid, or —NH—CH[(—CH<sub>2</sub>)<sub>5</sub>—COOH]—CO— for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties than the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and —(CH<sub>2</sub>)<sub>4</sub>COOH for the side chain of serine. These examples are not meant to be limiting.

In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner:

AA<sub>1</sub> is serine, glycine, alanine, cysteine or threonine;

AA<sub>2</sub> is alanine, threonine, glycine, cysteine or serine;

- AA<sub>3</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
- AA<sub>4</sub> is proline, leucine, valine, isoleucine or methionine;
- 5 AA<sub>5</sub> is tryptophan, alanine, phenylalanine, tyrosine or glycine;
- AA<sub>6</sub> is serine, glycine, alanine, cysteine or threonine;
- AA<sub>7</sub> is proline, leucine, valine, isoleucine or methionine;
- AA<sub>8</sub> is alanine, threonine, glycine, cysteine or serine;
- AA<sub>9</sub> is alanine, threonine, glycine, cysteine or serine;
- 10 AA<sub>10</sub> is leucine, isoleucine, methionine or valine;
- AA<sub>11</sub> is serine, glycine, alanine, cysteine or threonine;
- AA<sub>12</sub> is leucine, isoleucine, methionine or valine;
- AA<sub>13</sub> is leucine, isoleucine, methionine or valine;
- AA<sub>14</sub> is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or
- 15 unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- AA<sub>15</sub> is arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidino-butanoic acid
- AA<sub>16</sub> is proline, leucine, valine, isoleucine or methionine;
- AA<sub>17</sub> is serine, glycine, alanine, cysteine or threonine;
- 20 AA<sub>18</sub> is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- AA<sub>19</sub> is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- 25 AA<sub>20</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
- AA<sub>21</sub> is alanine, threonine, glycine, cysteine or serine;
- AA<sub>22</sub> is alanine, threonine, glycine, cysteine or serine;
- 30 AA<sub>23</sub> is histidine, serine, threonine, cysteine, lysine or ornithine;
- AA<sub>24</sub> is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

AA<sub>25</sub> is asparagine, aspartic acid,, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid; and

AA<sub>26</sub> is cysteine, histidine, serine, threonine, lysine or ornithine.

5           It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

          In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each  
10       codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal  
15       codons of the gene of interest are modified in this manner.

          When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino  
20       acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also  
25       be contained within a nucleotide sequence with a purification handle or be flanked by cleavage sites, such as proteolytic cleavage sites.

          Such polypeptide spacers may have from about 5 to about 40 amino acid residues. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are  
30       therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID

NO:26) and  $(\text{Gly}_4\text{-Ser})_y$ , wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

$(\text{Gly-Pro-Gly-Gly})_3$

SEQ ID NO:27 Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly

5  $(\text{Gly}_4\text{-Ser})_3$

SEQ ID NO:28 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

or  $(\text{Gly}_4\text{-Ser})_4$

SEQ ID NO:29 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser  
Gly Gly Gly Gly Ser.

10 Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

15 Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may  
20 be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest. Nucleotide sequences encoding other amino acid sequences which display pH  
25 sensitivity or chemical sensitivity may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase,  
30 factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the definition of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid.

In one embodiment of the present invention, a TAG sequence is linked to the gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using column purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:30. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:31. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID NO:32 and SEQ ID NO:33, respectively. However, it should be understood that any antigenic region may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:34) and a furin cleavage site. Constructs containing a TAG sequence are shown in Figures 2 and 3. In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:35.

#### Methods of Administering Transposon-Based Vectors

In addition to the transposon-based vectors described above, the present invention also includes methods of administering the transposon-based vectors to an animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic animal wherein a gene of interest is incorporated into cells other than the germline cells (somatic cells) of the animal. The transposon-based vectors of the present invention are administered to a reproductive organ of an animal via any method known to those of skill in the art. Preferred reproductive organs include an ovary, an oviduct, a mammary gland, and a fallopian tube.

In some embodiments, a transposon-based vector is directly administered to the reproductive organ. Direct administration encompasses injection into the organ, and in a preferred embodiment, a transposon-based vector is injected into the lumen of the oviduct, and more preferably, the lumen of the magnum or the infundibulum of



the oviduct. The transposon-based vectors may additionally or alternatively be placed in an artery supplying the reproductive organ. Administering the vectors to the artery supplying the ovary results in transfection of follicles and oocytes in the ovary to create a germline transgenic animal. Alternatively, supplying the vectors through an artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide for deposition in the egg white. In one embodiment, a transposon-based vector is administered into the lumen of the magnum or the infundibulum of the oviduct and to an artery supplying the oviduct. Indirect administration to the oviduct epithelium may occur through the cloaca. Direct administration into the mammary gland comprises introduction into the duct system of the mammary gland.

Administration of transposon-based vectors may occur in arteries supplying the ovary and or through direct intrathecal administration into the ovary through injection.

The transposon-based vectors may be administered in a single administration, multiple administrations, continuously, or intermittently. The transposon-based vectors may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, the transposon-based vector is administered to an animal in multiple administrations, each administration containing the vector and a different transfecting reagent.

The transposon-based vectors may be administered to the animal at any point during the lifetime of the animal, however, it is preferable that the vectors are administered prior to the animal reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken between approximately 14 and 16 weeks of age and to a quail between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is administered to an animal following an increase in proliferation of the oviduct epithelial cells and/or the tubular gland cells. Such an increase in proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the animal is an avian, the transposon-based vector is administered following an increase in proliferation of the oviduct epithelial cells and before the avian begins to produce egg white constituents.

In a preferred embodiment, the animal is an egg-laying animal, and more preferably, an avian. In one embodiment, between approximately 1 and 150  $\mu\text{g}$ , 1 and 100  $\mu\text{g}$ , 1 and 50  $\mu\text{g}$ , preferably between 1 and 20  $\mu\text{g}$ , and more preferably between 5 and 10  $\mu\text{g}$  of transposon-based vector DNA is administered to the oviduct of a bird.

5 Optimal ranges depend upon the type of bird and the bird's stage of sexual maturity. In a chicken, it is preferred that between approximately 1 and 100  $\mu\text{g}$ , or 5 and 50  $\mu\text{g}$  are administered. In a quail, it is preferred that between approximately 5 and 10  $\mu\text{g}$  are administered. Intraoviduct administration of the transposon-based vectors of the present invention result in incorporation of the gene of interest into the cells of the

10 oviduct as evidenced by a PCR positive signal in the oviduct tissue. In other embodiments, the transposon-based vector is administered to an artery that supplies the oviduct. These methods of administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

15 According to the present invention, the transposon-based vector is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed

20 include, but are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, TROJENE® and PLUS reagent from InVitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche; Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, the transposon-based vector is

25 administered to the oviduct and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:5, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

30

Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

#### Testing for and Breeding Animals Carrying the Transgene

Following administration of a transposon-based vector to an animal, DNA is extracted from the animal to confirm integration of the gene of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovarian route (including intraarterial administrations to arteries leading to the oviduct or ovary). Example 6 below describes isolation of a proinsulin/ENT TAG protein from a transgenic hen following ammonium sulfate precipitation and ion exchange chromatography. Figure 5 demonstrates successful administration of a transposon-based vector to a hen, successful integration of the gene of interest, successful production of a protein encoded by the gene of interest, and successful deposition of the protein in egg white produced by the transgenic hen.

Actual frequencies of integration may be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best features of in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question.

Breeding experiments are also conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment performed according to the present invention, each male bird was exposed to 2-3 different adult female birds for 3-4 days each. This procedure was continued with  
5 different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR. It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be  
10 performed to amplify gene copy numbers and protein output.

#### Production of Desired Proteins or Peptides in Egg White

In one embodiment, the transposon-based vectors of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These transposon-based vectors preferably contain one or more of an  
15 ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6, pp. 2305-12; A. Qiu et al., 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T.  
20 Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123. Examples of transposon-based vectors designed for production of a desired protein in an egg white are shown in Figures 2 and 3.

#### Production of Desired Proteins or Peptides in Egg Yolk

The present invention is particularly advantageous for production of  
25 recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins, lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein  
30 production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and re-folding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in

sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

In one embodiment of the present invention wherein germline transfection is  
5 obtained via intraovarian administration of the transposon-based vector, deposition of a desired protein into the egg yolk is accomplished in offspring by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This transposon-based vector can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a  
10 preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-  
15 species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-  
20 insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:22. In a preferred embodiment, the transposon-based vector contains a transposase gene  
25 operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

#### Isolation and Purification of Desired Protein or Peptide

For large-scale production of protein, an animal breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are  
30 obtained and identified through, for example, standard animal breeding procedures or PCR protocols.

Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis,



high performance liquid chromatography, immunoprecipitation and the like. Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

In one embodiment of the present invention, the animal in which the desired protein is produced is an egg-laying animal. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), Egg Science & Technology, Haworth Press, Binghamton, NY (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished by any of a number of polypeptide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, and the like, either alone or in combination. Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes electrophoresis. Successful isolation and purification is confirmed by standard analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

To facilitate or enable purification of a desired protein or peptide, transposon-based vectors may include one or more additional epitopes or domains. Such epitopes or domains include DNA sequences encoding enzymatic or chemical cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexa-histidine or other cationic amino acids; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art.

In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic

epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification, and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

5

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp  
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys  
Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys (SEQ ID NO:35)

10 The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:33). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:34). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:31) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

15 Isolation and purification of a desired protein is performed as follows:

1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.
2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).
- 20 3. Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin will eliminate virtually all extraneous egg white proteins except ovalbumin and the transgenic ovalbumin-TAG-desired protein.
4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.
- 25 5. Cleavage of the transgene product can be accomplished in at least one of two ways:
  - a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is
  - 30 added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.

b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.

c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.

6. A final separation of either of these two (5a or 5b) protein mixtures is made using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.

Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein. More specific methods for isolation using the TAG label is provided in the Examples. Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques applied are well documented in each of these settings.

In another representative embodiment, egg whites containing a protein of interest were pooled and separated, in any order, from the yolks and other egg constituents by methods known to one skilled in the art. A variety of such methods is described in manuals known in the art, such as *Egg Science & Technology*, W. Stadelman, *et al.* (Eds.), Haworth Press, Binghamton, NY (1995).

One non-limiting example of a method for isolating a desired peptide, polypeptide or protein from an egg white is as follows. It is to be understood that this method may be employed to isolate any desired peptide, polypeptide or protein from the eggs of transgenic animals of the present invention. This present example involved transgenes that used a portion of or the entire ovalbumin protein, or specific ovalbumin epitopes, as a carrier, linked to the protein of interest via the specified TAG sequence, or another affinity/cleavage sequence. The TAG sequence contains the hairpin loop epitope from HIV I followed by an enterokinase cleavage site.

First, the viscosity of the egg white was lowered by subjecting the egg white to low shear forces of 3140 cps (Tung et al., 1969). The resulting pourable solution

was then filtered to remove chalazae. An ammonium sulfate precipitation was then used to enrich the fraction of transgenic protein (see, for example, *Practical Protein Chemistry A Handbook* A. Darbre (Ed.), John Wiley & Sons Ltd., 1986). Other methods of crude fractionation known in the art are also used as needed. The supernatant of this separation was then fractionated using size-exclusion chromatography, further enriching the transgenic fusion protein fraction and eliminating the ammonium sulfate from the material. The fusion protein was isolated by anti-ovalbumin affinity chromatography (batch or column) using methods known to one skilled in the art. This step may capture native ovalbumin in addition to an ovalbumin-transgene fusion protein. After elution from the anti-ovalbumin affinity resin, the transgenic protein was specifically isolated using anti-gp41 affinity chromatography (batch or column) using methods known to one skilled in the art.

Cleavage of the transgene product from the carrier and the TAG sequences was accomplished in one of at least two ways:

1) The transgenic ovalbumin-TAG-transgene target protein was left attached to the gp41 affinity resin and the protease enterokinase was added. Cleavage of the transgene by enterokinase liberated the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence was retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), kept the transgene target protein together with enterokinase in solution. Enterokinase was recovered and reused.

2) Alternatively, enterokinase was immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-transgene target protein was eluted from the gp41 affinity chromatography resin and then applied to the protease resin. Protease action cleaved the ovalbumin-TAG sequence from the transgene target protein and left both entities in solution. The immobilized enterokinase resin was recharged and reused. The choice between these alternatives is made on a case-by case basis, depending upon the size and chemical composition of the transgene target protein.

A final separation of either of these two (process 1 or 2) protein mixtures was made using size exclusion chromatography, or enterokinase affinity chromatography. This step also allows for desalting, concentrating, buffer exchange and/or polishing, as needed.

It is believed that a typical chicken egg produced by a transgenic animal of the present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to the normal constituents of egg white (or possibly replacing a small fraction of the latter). In some embodiments, a chicken egg will contain between 50 and 75 mg of exogenous protein.

One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and reduce protein and then to cause the protein to re-fold into the preferred conformation. Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

#### Production of Protein or Peptide in Milk

In addition to methods of producing eggs containing transgenic proteins or peptides, the present invention encompasses methods for the production of milk containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal through the duct system. In one embodiment, the transposon-based vector contains a transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. Genes of interest can include, but are not limited to antiviral and antibacterial proteins and immunoglobulins. In other embodiments, a transposon-based vector is administered to the ovary of an animal and germline transformation is obtained. In these embodiments, offspring of the transfected animal express a gene of interest in the mammary gland under the control of a mammary gland-specific promoter.

The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.



## EXAMPLE 1

*IntraOviduct Administration of Transposon-Based Vectors*

Quail or chicken were selected for administration of the transposon-based vectors of the present invention. Feathers were removed from the area where surgery was performed and the area was cleansed and sterilized by rinsing it with ethanol (alcohol) and 0.5% chlorhexidine. Using the scalpel, a dorsolateral incision was made through the skin over the ovary approximately 2 cm in length. Using blunt scissors, a second incision was made through the muscle between the last two ribs to expose the oviduct beneath. A small animal retractor was used to spread the last two ribs, exposing the oviduct beneath. The oviduct was further exposed using retractors to pull the intestines to one side.

A delivery solution containing a transposon-based vector and SUPERFECT® was prepared fresh immediately before surgery. Specific ratios of vector and SUPERFECT® that were used in each experiment are provided in the Examples below. The delivery solution was warmed to room temperature prior to injection into the bird. Approximately 250-500 µl of the delivery solution was injected into the lumen of the magnum of the oviduct using a 1 cc syringe with a 27 gauge needle attached. The wound was closed and antibiotic cream liberally applied to the area surrounding the wound.

20

## EXAMPLE 2

*Preparation of Transposon-Based Vector pTnMod*

A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:3. The vector of SEQ ID NO:3, termed pTnMod, was constructed and its sequence verified.

This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:1) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic

30

polyA insures proper termination without read through to potential genes downstream.

The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACTTGATATTTTACATGATTCTCTTTACC 3' (SEQ ID NO:36) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTTCCCCAAAGCGTAACC 3' (SEQ ID NO:37, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:38) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:1) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:37) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:39), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe™ PCR System (which includes enzyme, buffers, dNTP's, MgCl<sub>2</sub> and PCR Enhancer, Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. A band corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting

the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the  
5 ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies, Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media  
10 (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates). These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide  
15 staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes  
20 and no further changes or mutations occurred due to PCR amplification. For sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

Once a clone was identified that contained the desired mATS in the correct  
25 orientation, primers CMVf-NgoM IV (5' TTGCCGGCATCAGATTGGCTAT (SEQ ID NO:40); underlined bases denote a NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGGTCACCGGGTCAATTCTTCAGCACCTGGTA (SEQ ID NO:41); underlined bases denote a BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the  
30 transposon in pTnLac. The PCR was conducted with FailSafe™ as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA were ligated together using a Stratagene T4 Ligase Kit and transformed into *E. coli* as described above.

Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:3) and includes the following components:

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptII sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

5       Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptII sk(20), corresponding to bp 924-235 of pBluescriptII sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites used in constructing the vector.

10       Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding  $\lambda$  DNA that is residual from pNK2859.

Base pairs 4645 - 5488 are non-coding DNA that is residual from pNK2859.

15       Base pairs 5489 - 7689 are from the pBluescriptII sk(-) base vector - (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptII sk(-).

Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

20       It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

25       All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500  $\mu$ L of PCR-grade water and stored at -20°C until used.

### EXAMPLE 3

#### *Transposon-Based Vector pTnMCS*

30       Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:2. The pTnMCS vector contains an avian optimized polyA sequence



operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 40 nucleotides that precede the A nucleotide string.

Bp 1 – 130 Remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) bp1-130

Bp 133 – 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy

5 Systems) bp 229-1873

Bp 1783 – 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316

Bp 2992 – 3344 Non coding DNA from vector pNK2859

Bp 3345 – 3387 Lambda DNA from pNK2859

Bp 3388 – 3457 70 bp of IS10 left from Tn10

10 Bp 3464 – 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718

Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673

15 Bp 3716 – 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235

Bp 4159 - 4228 70 bp of IS10 right from Tn10

Bp 4229 - 4270 Lambda DNA from pNK2859

Bp 4271 – 5114 Non-coding DNA from pNK2859

20 Bp 5115 - 7315 pBluescript sk (-) base vector (Stratagene, Inc.) bp 761-2961.

#### EXAMPLE 4

##### *Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/ProIns/PA) – Chicken*

25 A vector was designed to insert a human proinsulin coding sequence under the control of a chicken ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:42.

Base pairs 1 - 130 are a remainder of F1(-) ori of pBluescriptII sk(-)  
30 (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 133 – 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

Base pairs 1780 – 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

Base pairs 2988-2993 are two engineered stop codons.

Base pairs 2995 – 3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922- 2337 of pGWiz.

5 Base pairs 3415 – 3718 are non coding DNA that is residual from vector pNK2859.

Base pairs 3719 – 3761 are  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 – 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

10 Base pairs 3838 – 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

Base pairs 4050 - 4951 are a chicken ovalbumin promoter (including SDRE) that corresponds to base pairs 431-1332 of the chicken ovalbumin promoter in GenBank Accession Number J00895 M24999.

15 Base pairs 4958 - 6115 are a chicken ovalbumin signal sequence and ovalbumin gene that correspond to base pairs 66-1223 of GenBank Accession Number V00383.1. (The STOP codon being omitted).

Base pairs 6122 - 6271 are a TAG sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).

Base pairs 6272 – 6531 are a proinsulin gene.

20 Base pairs 6539 – 6891 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.

Base pairs 6897 - 7329 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

25 Base pairs 7335- 7404 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

Base pairs 7405 - 7446 are  $\lambda$  DNA that is residual from pNK2859.

Base pairs 7447 – 8311 are non coding DNA that is residual from pNK2859.

30 Base pairs 8312 - 10512 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

## EXAMPLE 5

*Transposon-Based Vector pTnMOD (CMV-CHOVg-ent-ProInsulin-synPA)*

A vector was designed to insert a proinsulin coding sequence under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:43.

- Bp 1 – 4045 from vector pTnMod, bp 1 - 4045
- Bp 4051 – 5695 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
- 10 Bp 5702 – 6855 Chicken ovalbumin gene taken from GenBank accession # V00383, bp 66-1219
- Bp 6862 - 7011 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- Bp 7012 – 7272 Human Proinsulin taken from GenBank accession # NM000207, bp 15 117-377
- Bp 7273 – 7317 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)
- Bp 7318 - 7670 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271
- 20 Bp 7672 – 11271 from cloning vector pTnMCS, bp 3716-7315

## EXAMPLE 6

*Transfection of Japanese Quail using a Transposon-based Vector containing a Proinsulin Gene via Oviduct Injections*

- 25 Two experiments were conducted in Japanese quail using transposon-based vectors containing either Oval promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A (SEQ ID NO:42) or CMV promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A (SEQ ID NO:43).

- 30 In the first experiment, the Oval promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A containing construct was injected into the lumen of the oviduct of sexually mature quail; three hens received 5  $\mu$ g at a 1:3 SUPERFECT® ratio and three received 10  $\mu$ g at a 1:3 SUPERFECT® ratio. As of the writing of the present application, at least one bird that received above-mentioned construct was producing human proinsulin in egg white (other birds remain to be tested). This

experiment indicates that 1) the DNA has been stable for at least 3 months; 2) protein levels are comparable to those observed with a constitutive promoter such as the CMV promoter; and 3) sexually mature birds can be injected and results obtained without the need for cell culture. It is estimated that each quail egg contains  
5 approximately 1.4 µg/ml of the proinsulin protein. It is also estimated that each transgenic chicken egg contains 50-75 mg of protein encoded by the gene of interest.

In the second experiment, the transposon-based vector containing CMV promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A was injected into the lumen of the oviduct of sexually immature Japanese quail. A total of 9 birds were  
10 injected. Of the 8 survivors, 3 produced human proinsulin in the white of their eggs for over 6 weeks. An ELISA assay described in detail below was developed to detect GP41 in the fusion peptide (Oval gene/GP41 Enterokinase TAG/Proinsulin) since the GP41 peptide sequence is unique and not found as part of normal egg white protein. In all ELISA assays, the same birds produced positive results and all controls worked  
15 as expected.

ELISA Procedure: Individual egg white samples were diluted in sodium carbonate buffer, pH 9.6, and added to individual wells of 96 well microtiter ELISA plates at a total volume of 0.1 ml. These plates were then allowed to coat overnight at 4°C. Prior to ELISA development, the plates were allowed warm to room  
20 temperature. Upon decanting the coating solutions and blotting away any excess, non-specific binding of antibodies was blocked by adding a solution of phosphate buffered saline (PBS), 1% (w/v) BSA, and 0.05% (v/v) Tween 20 and allowing it to incubate with shaking for a minimum of 45 minutes. This blocking solution was subsequently decanted and replaced with a solution of the primary antibody (Goat  
25 Anti-GP41 TAG) diluted in fresh PBS/BSA/Tween 20. After a two hour period of incubation with the primary antibody, each plate was washed with a solution of PBS and 0.05% Tween 20 in an automated plate washer to remove unbound antibody. Next, the secondary antibody, Rabbit anti-Goat Alkaline Phosphatase-conjugated, was diluted in PBS/BSA/Tween 20 and allowed to incubate 1 hour. The plates were then  
30 subjected to a second wash with PBS/Tween 20. Antigen was detected using a solution of *p*-Nitrophenyl Phosphate in Diethanolamine Substrate Buffer for Alkaline Phosphatase and measuring the absorbance at 30 minutes and 1 hour.

Additionally, a proinsulin fusion protein produced using a construct described above was isolated from egg white using ammonium sulfate precipitation and ion exchange chromatography. A pooled fraction of the isolated fusion protein was run on an SDS-PAGE gel shown in Figure 5, lanes 4 and 6. Lanes 1 and 10 of the gel contain molecular weight standards, lanes 2 and 8 contain non-transgenic chicken egg white, whereas lanes 3, 5, 7 and 9 are blank.

#### EXAMPLE 7

##### *Isolation of Human Proinsulin Using Anti-TAG Column Chromatography*

10 A HiTrap NHS-activated 1 mL column (Amersham) was charged with a 30 amino acid peptide that contained the gp-41 epitope containing gp-41's native disulfide bond that stabilizes the formation of the gp-41 hairpin loop. The 30 amino acid gp41 peptide is provided as SEQ ID NO:32. Approximately 10 mg of the peptide was dissolved in coupling buffer (0.2 M NaHCO<sub>3</sub>, 0.5 M NaCl, pH 8.3 and  
15 the ligand was circulated on the column for 2 hours at room temperature at 0.5 mL/minute. Excess active groups were then deactivated using 6 column volumes of 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3 and the column was washed alternately with 6 column volumes of acetate buffer (0.1 M acetate, 0.5 M NaCl, pH 4.0) and ethanolamine (above). The column was neutralized using 1 X PBS. The column was  
20 then washed with buffers to be used in affinity purification: 75 mM Tris, pH 8.0 and elution buffer, 100 mM glycine-HCl, 0.5 M NaCl, pH 2.7. Finally, the column was equilibrated in 75 mM Tris buffer, pH 8.0.

Antibodies to gp-41 were raised in goats by inoculation with the gp-41 peptide described above. More specifically, goats were inoculated, given a booster injection  
25 of the gp-41 peptide and blood samples were obtained by veinupuncture. Serum was harvested by centrifugation. Approximately 30 mL of goat serum was filtered to 0.45  $\mu$ M and passed over a TAG column at a rate of 0.5 mL/min. The column was washed with 75 mM Tris, pH 8.0 until absorbance at 280 nm reached a baseline. Three column volumes (3 mL) of elution buffer (100 mM glycine, 0.5 M NaCl, pH 2.7) was  
30 applied, followed by 75 mM Tris buffer, pH 8.0, all at a rate of 0.5 mL/min. One milliliter fractions were collected. Fractions were collected into 200  $\mu$ L 1 M Tris, pH 9.0 to neutralize acidic fractions as rapidly as possible. A large peak eluted from the column, coincident with the application the elution buffer. Fractions were pooled. Analysis by SDS-PAGE showed a high molecular weight species that separated into



two fragments under reducing condition, in keeping with the heavy and light chain structure of IgG.

Pooled antibody fractions were used to charge two 1 mL HiTrap NHS-activated columns, attached in series. Coupling was carried out in the same manner as that used for charging the TAG column.

#### Isolation of Ovalbumin-TAG-Proinsulin from Egg White

Egg white from quail and chickens treated by intra-oviduct injection of the CMV-ovalbumin-TAG-proinsulin construct were pooled. Viscosity was lowered by subjecting the allantoic fluid to successively finer pore sizes using negative pressure filtration, finishing with a 0.22  $\mu$ M pore size. Through the process, egg white was diluted approximately 1:16. The clarified sample was loaded on the Anti-TAG column and eluted in the same manner as described for the purification of the anti-TAG antibodies. A peak of absorbance at 280 nm, coincident with the application of the elution buffer, indicated that protein had been specifically eluted from the Anti-TAG column. Fractions containing the eluted peak were pooled for analysis.

The pooled fractions from the Anti-TAG affinity column were characterized by SDS-PAGE and western blot analysis. SDS-PAGE of the pooled fractions revealed a 60 kDal molecular weight band not present in control egg white fluid, consistent with the predicted molecular weight of the transgenic protein. Although some contaminating bands were observed, the 60 kDal species was greatly enriched compared to the other proteins. An aliquot of the pooled fractions was cleaved overnight at room temperature with the protease, enterokinase. SDS-PAGE analysis of the cleavage product, revealed a band not present in the uncut material that co-migrated with a commercial human proinsulin positive control. Western blot analysis showed specific binding to the 60 kDal species under non-reducing condition (which preserved the hairpin epitope of gp-41 by retaining the disulfide bond). Western analysis of the low molecular weight species that appeared upon cleavage with an anti-human proinsulin antibody, conclusively identified the cleaved fragment as human proinsulin.

### EXAMPLE 8

#### *Purification Procedures for Insulin*

##### I. ELISA data for egg characterization/identification

An ELISA was employed for the initial screening of eggs and, thereby, identification of hens producing positive eggs. With further modifications this procedure was used for the initial quantification of recombinant protein amounts. These procedures were aided by the successful purification of an initial stock of the recombinant proinsulin (RPI). This stock of protein is used in the development of a double antibody assay that increases the sensitivity and reduces the background in the assay. Subsequent identification of hens producing positive eggs obviate the need to screen each egg collected. Only periodic checks are needed to determine if production levels are consistent.

10 II. Egg White (EW) or Albumin Preparation

A. Clarification – Ovomucin precipitation

Eggs from hens positively identified as producing RPI are pooled for RPI purification. The initial purification step involved diluting the pool 1:1 with 100 mM Tris-HCl, pH 8 for a final concentration of 50 mM Tris-HCl. The pH of this solution was then adjusted to 6 and ovomucin was allowed to precipitate at 4°C for a minimum of 3hrs (preferably overnight) with constant stirring. The precipitated ovomucin was then pelleted and removed by centrifugation at 2400 x g. After collection of the RPI containing supernatant, the pH of this solution was readjusted to 8.

B. Filtration

20 To prepare the egg white for loading onto the column and, thereby, minimize the potential for clogging the columns during loading, the egg white solution was filtered to at least 0.45  $\mu$ m.

Initially, the ovomucin precipitated egg white solution was subjected to successive filtration steps with the pore size of the filtration membrane decreasing at each step. This procedure involved time and dilution of the egg white solution to reach 0.45  $\mu$ m filtration.

Amersham's hollow-fiber ultrafiltration apparatus was used to produced a column-ready solution filtered down to < 0.2  $\mu$ m with an undiluted starting solution. This approach minimized the time and the solution dilution needed to prepare the egg white solution for column loading.

30 III. Purification

A. Affinity Chromatography –

Using antibody with specificity to a synthetic peptide modeled after the enterokinase recognition site, initial purification schemes involved developing a one-step column purification procedure for the RPI.

Goats immunized with the synthetic Ent peptide were employed to produce anti-Ent Tag antiserum which was used in the egg screening ELISAs followed by antibody purification. The purified goat Anti-Ent Tag antibodies were covalently

bound to the matrix of HiTrap NHS-activated HP columns (Amersham) and subsequently used to specifically bind and purify the RPI.

5 An initial attempt was made to direct the first purification step against the ovalbumin portion of the recombinant protein using an antibody specific for the ovalbumin portion. The present purification scheme employed a combination of classical techniques such as ammonium sulfate precipitation, ion exchange, and gel filtration chromatography.

10 After the initial ovomucin precipitation, the egg white solution was subjected to protein precipitation using a 40% ammonium sulfate fractionation. The precipitated protein was subsequently collected via centrifugation and resuspended in 50 mM Tris-HCl, pH 8. The resuspended protein solution was dialyzed to remove residual  $(\text{NH}_4)_2\text{SO}_4$  or subjected to gel filtration to remove the  $(\text{NH}_4)_2\text{SO}_4$  and partially isolate the RPI from the remaining egg white protein. The RPI was further isolated via anion exchange chromatography using a 0 to 0.5M NaCl gradient in 50 mM Tris-HCl, pH 8. Two possible elution profiles were observed. One at approximately 25% of the 0.5 M NaCl gradient without  $(\text{NH}_4)_2\text{SO}_4$  precipitation. The second was observed at less than 16% gradient (approximately 7%) following 40%  $(\text{NH}_4)_2\text{SO}_4$  precipitation and a longer gradient. Fractions containing RPI were identified by SDS-PAGE analysis and pooled.

20 Three gel filtration columns, differing by column size and fractionation range, were employed in RPI purification and/or desalting: Superdex 75 10/300 GL, Hiload 26/60 Superdex 75, and Hiload 26/60 Superdex 200. Using these individual columns at different steps in the purification scheme increased the efficiency of the process. Fractions containing RPI were identified by SDS-PAGE analysis and pooled.

25 Cleavage of the RPI Enterokinase recognition site was accomplished using purified enterokinase from Sigma. Enterokinase, 0.004 Unit/ $\mu\text{l}$  per reaction, was applied to the pooled and, if necessary, concentrated protein solution. The digestion reaction was incubated at room temperature (up to 30°C in a rolling hybridization oven) for a minimum of 16 h and in some cases up to 48 hrs of incubation. The digestion efficiency was followed using 16.5% Tris-Tricine SDS-PAGE peptide gels. All gel staining utilized Simply Blue Coomassie Staining Solutions. Free Proinsulin was observed on gels after digestion.

30 A subsequent gel filtration separation was employed to obtain purified Proinsulin, and to remove the remaining Ovalbumin portion of the RPI and residual native EW proteins. Select steps in the purification process were analyzed using the 2-dimensional Beckman Coulter ProteomeLab PF2D Protein Fractionation System.

## EXAMPLE 9

*Optimization of Intra-oviduct and Intra-ovarian Arterial Injections*

Overall transfection rates of oviduct cells in a flock of chicken or quail hens are enhanced by synchronizing the development of the oviduct and ovary within the flock. When the development of the oviducts and ovaries are uniform across a group of hens and when the stage of oviduct and ovarian development can be determined or predicted, timing of injections is optimized to transfect the greatest number of cells. Accordingly, oviduct development is synchronized as described below to ensure that a large and uniform proportion of oviduct secretory cells are transfected with the gene of interest.

Hens are treated with estradiol to stimulate oviduct maturation as described in Oka and Schimke (T. Oka and RT Schimke, J. Cell Biol., 41, 816 (1969)), Palmiter, Christensen and Schimke (J Biol. Chem. 245(4):833-845, 1970). Specifically, repeated daily injections of 1 mg estradiol benzoate are performed sometime before the onset of sexual maturation, a period ranging from 1 – 14 weeks of age. After a stimulation period sufficient to maximize development of the oviduct, hormone treatment is withdrawn thereby causing regression in oviduct secretory cell size but not cell number. At an optimum time after hormone withdrawal, the lumens of the oviducts of treated hens are injected with the transposon-based vector. Hens are subjected to additional estrogen stimulation after an optimized time during which the transposon-based vector is taken up into oviduct secretory cells. Re-stimulation by estrogen activates transposon expression, causing the integration of the gene of interest into the host genome. Estrogen stimulation is then withdrawn and hens continue normal sexual development. If a developmentally regulated promoter such as the ovalbumin promoter is used, expression of the transposon-based vector initiates in the oviduct at the time of sexual maturation. Intra-ovarian artery injection during this window allows for high and uniform transfection efficiencies of ovarian follicles to produce germ-line transfections and possibly oviduct expression.

Other means are also used to synchronize the development, or regression, of the oviduct and ovary to allow high and uniform transfection efficiencies. Alterations of lighting and/or feed regimens, for example, cause hens to 'molt' during which time the oviduct and ovary regress. Molting is used to synchronize hens for transfection, and may be used in conjunction with other hormonal methods to control regression and/or development of the oviduct and ovary.

## EXAMPLE 10

*Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/ProIns/PA) – Quail*

- 5           A vector is designed for inserting a proinsulin gene under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:44.

10           Base pairs 1 -130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

          Base pairs 133 – 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

          Base pairs 1780 – 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

15           Base pairs 2988-2993 are an engineered stop codon.

          Base pairs 2995 – 3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922- 2337 of pGWiz.

          Base pairs 3415 – 3718 are non coding DNA that is residual from vector pNK2859.

20           Base pairs 3719 – 3761 are  $\lambda$  DNA that is residual from pNK2859.

          Base pairs 3762 – 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

          Base pairs 3838 – 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

25           Base pairs 4050 - 4938 are the Japanese quail ovalbumin promoter (including SDRE, steroid-dependent response element). The Japanese quail ovalbumin promoter was isolated by its high degree of homology to the chicken ovalbumin promoter (GenBank accession number J00895 M24999, base pairs 431-1332). Some deletions were noted in the quail sequence, as compared to the chicken sequence.

30           Base pairs 4945 - 6092 are a quail ovalbumin signal sequence and ovalbumin gene that corresponds to base pairs 54 – 1201 of GenBank accession number X53964.1. (The STOP codon being omitted).

          Base pairs 6093 - 6246 are a TAG sequence containing a gp41 hairpin loop from HIV I an enterokinase cleavage site and a spacer (synthetic).



Base pairs 6247 – 6507 are a proinsulin gene.

Base pairs 6514 – 6866 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.

Base pairs 6867 - 7303 are a multiple cloning site from pBlueScriptII sk(-)  
5 corresponding to base pairs 667-235 of pBluescriptII sk(-).

Base pairs 7304- 7379 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

Base pairs 7380 - 7421 are  $\lambda$  DNA that is residual from pNK2859.

Base pairs 7422 – 8286 are non coding DNA that is residual from pNK2859.

10 Base pairs 8287 - 10487 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences  
15 in the above construct represent restriction site remnants.

#### EXAMPLE 11

##### *Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/p146/PA) – Chicken*

20 A vector was designed for inserting a p146 gene under the control of a chicken ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird. The vector sequence is provided below as SEQ ID NO:45.

Base pairs 1 - 130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

25 Base pairs 133 – 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

Base pairs 1780 – 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

Base pairs 2988-2993 are an engineered stop codon.

30 Base pairs 2995 – 3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922- 2337 of pGWiz.

Base pairs 3415 – 3718 are non coding DNA that is residual from vector pNK2859.

Base pairs 3719 – 3761 are  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 – 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

Base pairs 3838 – 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

5        Base pairs 4050 - 4951 are a chicken ovalbumin promoter (including SDRE, steroid-dependent response element) that corresponds to base pairs 431-1332 of the chicken ovalbumin promoter in GenBank Accession Number J00895 M24999.

Base pairs 4958 - 6115 are a chicken ovalbumin signal sequence and Ovalbumin gene that correspond to base pairs 66-1223 of GenBank Accession  
10    Number V00383.1 (The STOP codon being omitted).

Base pairs 6122 - 6271 are a TAG sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).

Base pairs 6272 – 6316 are a p146 sequence (synthetic) with 2 added stop codons.

15        Base pairs 6324 – 6676 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.

Base pairs 6682 - 7114 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

Base pairs 7120- 7189 are the 70 base pairs of the right insertion sequence  
20    (IS10) recognized by the transposon Tn10.

Base pairs 7190 - 7231 are  $\lambda$  DNA that is residual from pNK2859.

Base pairs 7232 – 8096 are non coding DNA that is residual from pNK2859.

Base pairs 8097 - 10297 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

25

It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

30

## EXAMPLE 12

### *Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/p146/PA) – Quail*

A vector was designed for inserting a p146 gene under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird. The vector sequence is given below as SEQ ID NO:46.

Base pairs 1 - 130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 133 - 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

5       Base pairs 1780 - 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

Base pairs 2988-2993 are an engineered stop codon.

Base pairs 2995 - 3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922-2337 of pGWiz.

10       Base pairs 3415 - 3718 are non coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are  $\lambda$  DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

15       Base pairs 3838 - 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

Base pairs 4050 - 4938 are the Japanese quail ovalbumin promoter (including SDRE, steroid-dependent response element). The Japanese quail ovalbumin promoter was isolated by its high degree of homology to the chicken ovalbumin promoter  
20 (GenBank accession number J00895 M24999, base pairs 431-1332).

Bp 4945 - 6092 are a quail ovalbumin signal sequence and ovalbumin gene that corresponds to base pairs 54 - 1201 of GenBank accession number X53964.1. (The STOP codon being omitted).

Base pairs 6097 - 6246 are a TAG sequence containing a gp41 hairpin loop  
25 from HIV I, an enterokinase cleavage site and a spacer (synthetic).

Base pairs 6247 - 6291 are a p146 sequence (synthetic) with 2 added stop codons.

Base pairs 6299 - 6651 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.

30       Base pairs 6657 - 7089 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

Base pairs 7095- 7164 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

Base pairs 7165 - 7206 are  $\lambda$  DNA that is residual from pNK2859.

Base pairs 7207 – 8071 are non coding DNA that is residual from pNK2859.

Base pairs 8072 - 10272 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

- 5           It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

#### EXAMPLE 13

##### 10   *Additional Transposon-Based Vectors for Administration to an Animal*

- The following example provides a description of various transposon-based vectors of the present invention and several constructs that have been made for insertion into the transposon-based vectors of the present invention. These examples are not meant to be limiting in any way. The constructs for insertion into a transposon-based vector are provided in a cloning vector pTnMCS or pTnMod, both described above.

##### pTnMCS (CMV-CHOVg-ent-ProInsulin-synPA) (SEQ ID NO:47)

Bp 1 – 3670 from vector PTnMCS, bp 1 - 3670

- 20   Bp 3676 – 5320 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems), bp 230-1864

Bp 5327 – 6480 Chicken ovalbumin gene taken from GenBank accession # V00383, bp 66-1219

- Bp 6487 - 6636 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 6637 – 6897 Human Proinsulin taken from GenBank accession # NM000207, bp 117-377

Bp 6898 – 6942 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

- 30   Bp 6943 - 7295 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

Bp 7296 – 10895 from cloning vector pTnMCS, bp 3716-7315

pTnMCS (CMV-prepro-ent-ProInsulin-synPA)

Bp 1 – 3670 from vector PTnMCS, bp 1 - 3670

Bp 3676 – 5320 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems), bp 230-1864

5 Bp 5326 - 5496 Capsite/prepro taken from GenBank accession # X07404, bp 563–733

Bp 5504 - 5652 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 5653 – 5913 Human Proinsulin taken from GenBank accession # NM000207, bp 117-377

10 Bp 5914 – 5958 Spacer DNA, derived as an artifact from the cloning vectors pTOPO

Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

Bp 5959-6310 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

Bp 6313–9912 from cloning vector pTnMCS, bp 3716-7315

15

pTnMCS(Chicken OVep+OVg'+ENT+proins+syn polyA)

Bp 1–3670 from vector pTnMCS, bp 1 - 3670

Bp 3676–4350 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1 bp 1-675

20 Bp 4357–5692 Chicken Ovalbumin promoter taken from GenBank accession # J00895M24999 bp 1-1336

Bp 5699–6917 Chicken Ovalbumin gene from GenBank Accession # V00383.1 bp 2-1220. (This sequence includes the 5'UTR, containing putative cap site, bp 5699-5762.)

25 Bp 6924-7073 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 7074-7334 Human proinsulin GenBank Accession # NM000207 bp 117-377

Bp 7335-7379 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)

30 Bp 7380-7731 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271

Bp 7733–11332 from vector pTnMCS, bp 3716 - 7315



pTnMCS(Chicken OVep+prepro+ENT+proins+syn polyA)

Bp 1 – 3670 from cloning vector pTnMCS, bp 1 - 3670

Bp 3676 – 4350 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1 bp 1-675

- 5 Bp 4357 – 5692 Chicken Ovalbumin promoter taken from GenBank accession # J00895-M24999 bp 1-1336

Bp 5699–5869 Cecropin cap site and prepro, Genbank accession # X07404 bp 563-733

- 10 Bp 5876 - 6025 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 6026 - 6286 Human proinsulin GenBank Accession # NM000207 bp 117-377

Bp 6287 - 6331 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)

- 15 Bp 6332 - 6683 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271

Bp 6685 – 10284 from cloning vector pTnMCS, bp 3716 - 7315

pTnMCS(Quail OVep+OVg'+ENT+proins+syn polyA)

Bp 1 – 3670 from cloning vector pTnMCS, bp 1 - 3670

- 20 Bp 3676 – 4333 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession # S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)

- 25 Bp 4340 – 5705 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)

- 30 Bp 5712 – 6910 Quail Ovalbumin gene, EMBL accession # X53964, bp 1-1199. (This sequence includes the 5'UTR, containing putative cap site bp 5712-5764.)

Bp 6917 - 7066 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 7067 - 7327 Human proinsulin GenBank Accession # NM000207 bp 117-377

- Bp 7328 - 7372 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- Bp 7373 - 7724 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271
- 5 Bp 7726 - 11325 from cloning vector pTnMCS, bp 3716 - 7315
- pTnMCS(Quail OVep+prepro+ENT+proins+syn polyA)
- Bp 1 - 3670 from cloning vector pTnMCS, bp 1 - 3670
- Bp 3676 - 4333 Quail Ovalbumin enhancer: 658 bp sequence, amplified from quail
- 10 genomic DNA, roughly equivalent to the far- upstream chicken ovalbumin enhancer, GenBank accession #S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
- Bp 4340 - 5705 Quail Ovalbumin promoter: 1366 bp sequence, amplified from quail
- 15 genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
- Bp 5712-5882 Cecropin cap site and prepro, Genbank accession # X07404 bp 563-
- 20 733
- Bp 5889 - 6038 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- Bp 6039 - 6299 Human proinsulin GenBank Accession # NM000207 bp 117-377
- Bp 6300 - 6344 Spacer DNA, derived as an artifact from the cloning vectors pTOPO
- 25 Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- Bp 6345 - 6696 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271
- Bp 6698 - 10297 from cloning vector pTnMCS, bp 3716 - 7315.
- 30 pTnMOD (CMV-prepro-ent-proins-synPA)
- Bp 1 - 4045 from vector PTnMCS, bp 1 - 4045
- Bp 4051 - 5695 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
- Bp 5701-5871 Capsite/prepro taken from GenBank accession # X07404, bp 563-733

- Bp 5879 - 6027 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- Bp 6028-6288 Human Proinsulin taken from GenBank accession # NM000207, bp 117-377
- 5 Bp 6289 - 6333 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)
- Bp 6334 - 6685 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271
- Bp 6687 -10286 from cloning vector pTnMCS, bp 3716-7315
- 10 pTnMOD(Chicken OVep+OVg'+ENT+proins+syn polyA)
- Bp 1 - 4045 from cloning vector pTnMod, bp 1 - 4045
- Bp 4051 - 4725 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1 bp 1-675
- 15 Bp 4732 - 6067 Chicken Ovalbumin promoter taken from GenBank accession # J00895-M24999 bp 1-1336
- Bp 6074 - 7292 Chicken Ovalbumin gene from GenBank Accession # V00383.1 bp 2-1220. (This sequence includes the 5'UTR, containing putative cap site bp 6074-6137.)
- 20 Bp 7299 - 7448 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- Bp 7449 - 7709 Human proinsulin GenBank Accession # NM000207 bp 117-377
- Bp 7710 - 7754 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- 25 Bp 7755 - 8106 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271
- Bp 8108 - 11707 from cloning vector pTnMod, bp 3716 - 7315
- pTnMOD(Chicken OVep+prepro+ENT+proins+syn polyA)
- 30 Bp 1 - 4045 from cloning vector pTnMCS, bp 1 - 4045
- Bp 4051 - 4725 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1 bp 1-675
- Bp 4732 - 6067 Chicken Ovalbumin promoter taken from GenBank accession # J00895-M24999 bp 1-1336

- Bp 6074–6244 Cecropin cap site and prepro, Genbank accession # X07404 bp 563-733
- Bp 6251 - 6400 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- 5 Bp 6401 - 6661 Human proinsulin GenBank Accession # NM000207 bp 117-377  
Bp 6662 - 6706 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)  
Bp 6707 - 7058 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271
- 10 Bp 7060 – 10659 from cloning vector pTnMCS, bp 3716 - 7315
- pTnMOD(Quail OVep+OVg'+ENT+proins+syn polyA)
- Bp 1 – 4045 from cloning vector pTnMCS, bp 1 - 4045
- Bp 4051 – 4708 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house
- 15 from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession # S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
- Bp 4715 – 6080 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house
- 20 from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
- Bp 6087 – 7285 Quail Ovalbumin gene, EMBL accession # X53964, bp 1-1199. (This
- 25 sequence includes the 5'UTR, containing putative cap site bp 6087-6139.)
- Bp 7292 - 7441 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- Bp 7442 - 7702 Human proinsulin GenBank Accession # NM000207 bp 117-377
- Bp 7703 - 7747 Spacer DNA, derived as an artifact from the cloning vectors pTOPO
- 30 Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- Bp 7748 - 8099 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271
- Bp 8101 – 11700 from cloning vector pTnMCS, bp 3716 - 7315

pTnMOD(Quail OVep+prepro+ENT+proins+syn polyA)

Bp 1 – 4045 from cloning vector pTnMCS, bp 1 - 4045

Bp 4051 – 4708 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession #S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)

Bp 4715 – 6080 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)

Bp 6087–6257 Cecropin cap site and Prepro, Genbank accession # X07404 bp 563-733

Bp 6264 - 6413 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 6414 - 6674 Human proinsulin GenBank Accession # NM000207 bp 117-377

Bp 6675 - 6719 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)

Bp 6720 - 7071 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271

Bp 7073 – 10672 from cloning vector pTnMCS, bp 3716 - 7315

pTnMOD (CMV-prepro-ent-hGH-CPA)

Bp 1–4045 from vector PTnMOD, bp 1 - 4045

Bp 4051–5694 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1873

Bp 5701-5871 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

Bp 5878-6012 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 6013–6666 Human growth hormone taken from GenBank accession # V00519, bp 1-654

Bp 6673–7080 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058



Bp 7082–10681 from cloning vector pTnMOD, bp 4091–7690

pTnMCS (CHOVep-prepro-ent-hGH-CPA)

Bp 1–3670 from vector PTnMCS, bp 1–3670

- 5 Bp 3676–4350 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1–675

Bp 4357–5692 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1–1336

Bp 5699–5869 Capsite/Prepro taken from GenBank accession # X07404, bp 563–733

- 10 Bp 5876–6010 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 6011–6664 Human growth hormone taken from GenBank accession # V00519, bp 1–654

- 15 Bp 6671–7078 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651–11058

Bp 7080–10679 from cloning vector pTnMCS, bp 3716–7315

pTnMCS (CMV-prepro-ent-hGH-CPA)

Bp 1 – 3670 from vector PTnMCS, bp 1 - 3670

- 20 Bp 3676–5319 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230–1873

Bp 5326–5496 Capsite/Prepro taken from GenBank accession # X07404, bp 563 – 733

Bp 5503–5637 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

- 25 Bp 5638–6291 Human growth hormone taken from GenBank accession # V00519, bp 1–654

Bp 6298–6705 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651–11058

Bp 6707–10306 from cloning vector pTnMCS, bp 3716–7315

30

pTnMOD (CHOVep-prepro-ent-hGH-CPA)

Bp 1–4045 from vector PTnMOD, bp 1–4045

Bp 4051–4725 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1–675

- Bp 4732-6067 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
- Bp 6074-6244 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733
- Bp 6251-6385 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- 5 Bp 6386-7039 Human growth hormone taken from GenBank accession # V00519, bp 1-654
- Bp 7046-7453 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
- 10 Bp 7455-11054 from cloning vector pTnMOD, bp 4091-7690

PTnMod(CMV/Transposase/ChickOvep/prepro/ProteinA/ConpolyA)

- BP 1-130 remainder of F1 (-) ori of pBluescriptII sk(-) (Stragagene) bp 1-130.
- BP 133-1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873.
- 15 BP 1780-2987 Transposase, modified from Tn10 (GenBank #J01829).
- BP 2988-2993 Engineered DOUBLE stop codon.
- BP 2994-3343 non coding DNA from vector pNK2859.
- BP 3344-3386 Lambda DNA from pNK2859.
- 20 BP 3387-3456 70bp of IS10 left from Tn10.
- BP 3457-3674 multiple cloning site from pBluescriptII sk(-) bp 924-707.
- BP 3675-5691 Chicken Ovalbumin enhancer plus promoter from a Topo Clone 10 maxi 040303 (5' XmaI, 3' BamHI)
- BP 5698-5865 prepro with Cap site amplified from cecropin of pMON200 GenBank # X07404 (5'BamHI, 3'KpnI)
- 25 BP 5872-7338 Protein A gene from GenBank# J01786, mature peptide bp 292-1755 (5'KpnI, 3'SacII)
- BP 7345-7752 ConPolyA from Chicken conalbumin polyA from GenBank # Y00407 bp 10651-11058. (5'SacII, 3'XhoI)
- 30 BP 7753-8195 multiple cloning site from pBluescriptII sk(-) bp 677-235.
- BP 8196-8265 70 bp of IS10 left from Tn10.
- BP 8266-8307 Lamda DNA from pNK2859
- BP 8308-9151 noncoding DNA from pNK2859
- BP 9152-11352 pBluescriptII sk(-) base vector (Stratagene, INC.) bp 761-2961

All patents, publications and abstracts cited above are incorporated herein by reference in their entirety. It should be understood that the foregoing relates only to preferred embodiments of the present invention and that numerous modifications or  
5 alterations may be made therein without departing from the spirit and the scope of the present invention as defined in the following claims.

## CLAIMS

We Claim:

- 5 1. A method of producing a transgenic animal comprising, administering to an oviduct or an ovary of an animal a composition comprising a transposon-based vector.
- 10 2. The method of claim 1, wherein the composition is injected into an artery leading to the oviduct or the ovary.
3. The method of claim 1, wherein the composition is injected into a lumen of the oviduct.
- 15 4. The method of claim 1, wherein the composition further comprises a transfection reagent.
5. The method of claim 1, wherein the transposon-based vector comprises:
  - 20 a) a transposase gene operably linked to a first promoter, the transposase gene encoding for a transposase; and
  - b) one or more genes of interest operably-linked to one or more additional promoters; wherein the one or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by the transposase, and wherein the first promoter comprises a  
25 modified Kozak sequence comprising ACCATG (SEQ ID NO:1).
6. The method of claim 5, wherein one to twenty codons at a beginning of the transposase gene are modified by changing a nucleotide at a third base  
30 position of the codon to an adenine or thymine without modifying the amino acid encoded by the codon.
7. The method of claim 1, wherein the transposon-based vector comprises:
  - 35 a) a transposase gene operably linked to a first promoter and an avian optimized polyA sequence, the transposase gene encoding for a transposase; and
  - b) one or more genes of interest operably-linked to one or more additional promoters;

c) wherein the one or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by the transposase.

5           8.     The method of claim 7, wherein the first promoter is a constitutive promoter.

10           9.     The method of claim 7, wherein the first promoter is an oviduct-specific promoter selected from the group consisting of ovalbumin, ovotransferrin, ovomucoid, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin.

15           10.    The method of claim 7, wherein the one or more gene of interest is operably-linked to a second promoter.

20           11.    The method of claim 10, wherein the second promoter is an oviduct-specific promoter selected from the group consisting of ovalbumin, ovotransferrin, ovomucoid, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin.

25           12.    The method of claim 7, wherein the transposon-based vector further comprises an egg directing sequence or an enhancer operably-linked to the one or more genes of interest.

30           13.    A method of increasing expression of a polynucleotide sequence in an avian comprising, administering to the avian a composition comprising the polynucleotide sequence operably-linked to an avian optimized polyA sequence.

35           14.    A transgenic animal made with the method of Claim 1.

15.    The transgenic animal of Claim 14, wherein the animal is a bird.

16.    An egg produced by the transgenic bird of Claim 15.

17.    The egg of Claim 16, wherein the egg contains a protein, a polypeptide or a peptide encoded by the transposon-based vector.



18. A method of producing proteins, polypeptides or peptides comprising:  
a) administering to an oviduct of an egg-laying animal a composition comprising a transposon-based vector, wherein transposon-based vector comprises a transposase gene operably linked to a first promoter, the transposase gene encoding for a transposase, and one or more genes of interest operably-linked to one or more additional promoters; and,  
b) permitting the one or more genes of interest to be expressed into a protein, a polypeptide or a peptide.
19. The method of Claim 18, further comprising  
a) collecting an egg from the egg-laying animal;  
b) harvesting egg white containing the protein, the polypeptide or the peptide; and,  
c) purifying the protein, the polypeptide or the peptide.
20. The method of any of the preceding claims, wherein the animal is a bird.
21. The method of any of the preceding claims, wherein the animal is a poultry bird.
22. The method of any of the preceding claims, wherein the transposase is a Tn10 transposase.
23. A method of producing a transgenic mammal comprising, administering to an intramammary duct system of a mammal a composition comprising a transposon-based vector.
24. A transgenic mammal made with the method of Claim 23.
25. Milk produced by the transgenic mammal of Claim 24.
26. The milk of Claim 25, wherein the milk contains a protein, a polypeptide or a peptide encoded by the transposon-based vector.
27. A method of producing proteins, polypeptides or peptides comprising:  
a) administering to an intramammary duct system of a mammal a composition comprising a transposon-based vector, wherein transposon-based

vector comprises a transposase gene operably linked to a first promoter, the transposase gene encoding for a transposase, and one or more genes of interest operably-linked to one or more additional promoters; and,

- 5           b)     permitting the one or more genes of interest to be expressed into a protein, a polypeptide or a peptide.

28.   The method of Claim 27, further comprising

- a)     collecting milk from the mammal; and,  
b)     purifying the protein, the polypeptide or the peptide.

10

FIGURE 1



FIGURE 2



FIGURE 3

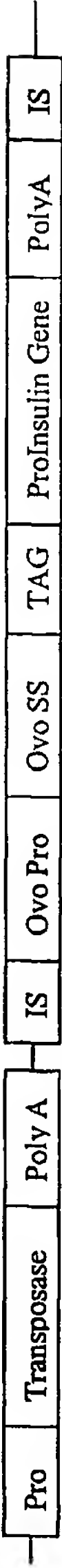


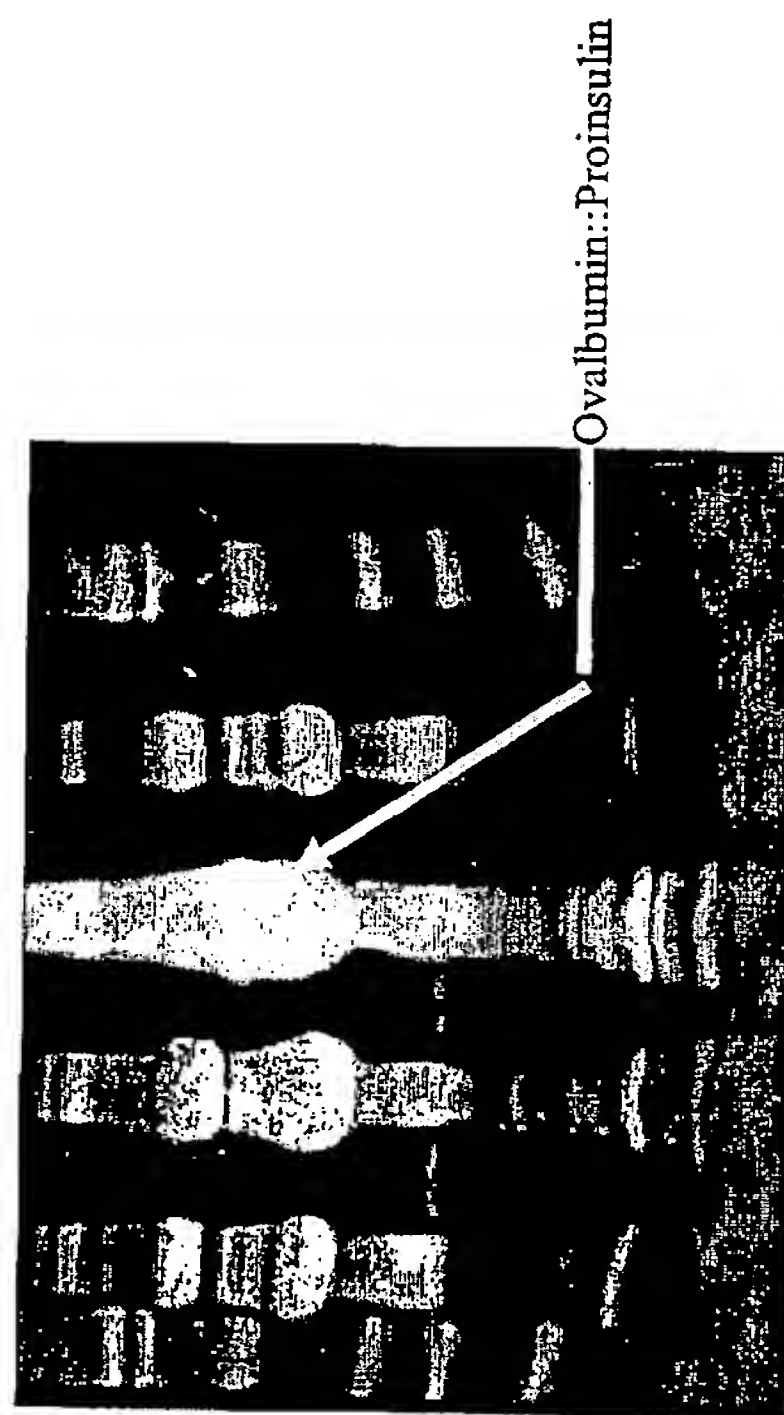


FIGURE 4

IS	Tet; Pro	Ovgen	Pro	Ovotrans	Pro	Ovomucin	IS
----	----------	-------	-----	----------	-----	----------	----

5/5

FIGURE 5



## Appendix A

SEQ ID NO:1 (modified Kozak sequence)  
ACCATG

5

SEQ ID NO:2 (pTnMCS)

1 ctgacgcgcc ctgtacggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga  
61 ccgtacact tgccagcgcc ctacgccccg ctcctttcgc tttcttccct tcctttctcg  
10 121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa  
181 tatgtacatt tataattggct catgtccaac attaccgcca tgttgacatt gattattgac  
241 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg  
301 cgttacataa cttacggtaa atggccccgc tggctgaccg cccaacgacc cccgcccatt  
361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
15 421 atgggtggag tatttacggg aaactgcccc cttggcagta catcaagtgt atcatatgcc  
481 aagtagcccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
541 catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tcgctattac  
601 catggtgatg cggtttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg  
661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg  
20 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt  
781 acgggtggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg  
841 ccattccacg tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcgcccg  
901 ggaacgggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgctatag  
961 actctatagg cacaccctt tggctcttat gcatgctata ctgttttttg cttggggcct  
25 1021 atacaccccc gcttccttat gctatagggt atggtatagc ttagcctata ggtgtgggtt  
1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac  
1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac  
1201 tgacacggac tctgtatttt tacaggatgg ggtcccatth attatttaca aattcacata  
1261 tacaacaacg ccgtcccccg tgcccgagc ttttattaaa catagcgtgg gatctccacg  
30 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca  
1381 tccgagccct ggtcccatgc ctccagcggc tcatggctgc tcggcagctc cttgctccta  
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag  
1501 gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac  
1561 gcagatggaa gacttaaggg agcggcagaa gaagatgcag gcagctgagt tgtgtattc  
35 1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cgggtggaggg cagtgtagtc  
1681 tgagcagtag tegtgtctgc cgcgcgcgcc accagacata atagctgaca gactaacaga  
1741 ctgttccttt ccatgggtct tttctgcagt cacogtcgga ccatgtgcca actcgatatt  
1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta  
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga acttggccgt  
40 1921 aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt  
1981 aatcgtcacc tccacaaaga gcgactcgt gtataccgtt ggcattgctag ctttatctgt  
2041 tcgggcaata cgtgcccatt tgtactgtt gactggctgt atattcgtga gcaaaaacga  
2101 cttatgggat tgcgagcttc agtcgcacta cacggctcgt ctgttactct ttatgagaaa  
2161 gcgttccgcg tttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt  
45 2221 gcgagcattc taccgagtaa caccacaccg ctcatgttca gtgatgctgg ctttaaagtg  
2281 ccatgggtata aatccgttga gaagctgggt tggtagctgg taagtgcagt aagaggaaaa  
2341 gtacaatatg cagacctagg agcggaaaac tggaaaaccta tcagcaactt acatgatatg  
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca  
2461 tgccaaattc tattgtataa atctcgtct aaaggccgaa aaaatcagcg ctgcacacgg  
50 2521 actcattgtc accacccgtc acctaaaatc tactcagcgt cggcaaagga gccatgggtt  
2581 ctacgaacta acttacctgt tgaaattcga acacccaaac aacttgtaa tatctattcg  
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc  
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg  
2761 atgcttcaac taacatgttg gcttgccggc gttcatgctc agaaaacaagg ttgggacaag  
55 2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg  
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc  
2941 ctactagctc aaaatttatt cacacatggg tacgcttttg ggaaattatg aggggatcgc  
3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaagggtgct tttatcatca  
3061 ctttaaaaaa aaaaaacaat tactcagtgc ctgttataag cagcaattaa ttatgattga  
60 3121 tgcctacatc acaacaaaaa ctgatttaac aaatgggttg tctgccttag aaagtatatt  
3181 tgaacattat cttgattata ttattgataa taataaaaaa cttatcccta tccaagaagt  
3241 gatgcctatc attgggttga atgaacttga aaaaaattag cettgaatac attactggta  
3301 aggtaaaacgc cattgtcagc aaattgatcc aagagaacca acttaaaagt ttctgacgg  
3361 aatgttaatt ctggttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaaac  
65 3421 attagtttaa ggtggatata catcttgtca tatgatcccc gtaatgtgag ttagctcact  
3481 cattaggcac cccaggcttt acactttatg cttccggctc gtatgtttgt tggaaattgtg

3541 agcggataac aatttcacac aggaacacagc tatgaccatg attacgcca ggcgcgaatt  
3601 aaccctcact aaaggggaaca aaagctggag ctccaccgcg gtggcgcccg ctctagaact  
3661 agtggatccc ccgggctgca ggaattcgat atcaagctta tcgataccgc tgacctcgag  
3721 gggggggcccg gtaccaatt cgccctatag tgagtcgtat tacgcgcgct cactggccgt  
5 3781 cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgacgc  
3841 acatccccct ttcgccagct ggcgtaatag cgaagaggcc cgcaccgac gcccttccca  
3901 acagttgcgc agcctgaatg gcgaatggaa attgtaagcg ttaatatatt gttaaaattc  
3961 gcgttaaatt tttgttaaat cagctcattt ttaaccaat aggcggaat cggcaaaatc  
10 4021 ccttataaat caaaagaata gaccgagata ggggttgagt ttgttccagt ttggaacaag  
4081 agtccactat taaagaacgt ggactccaac gtcaaaggcg gaaaaaccgt ctatcaggcg  
4141 gatggccac tactccggga tcatatgaca agatgtgtat ccaccttaac ttaatgattt  
4201 ttacaaaat cattagggga tcatcagtg ctcagggtca acgagaatta acattccgtc  
4261 aggaagcctt atgatgatga tgtgcttaaa aacttactca atggctgggt atgcataatc  
4321 caatacatgc gaaaaacctt aaagagcctg ccgataaaaa aggccaatat attgctattt  
15 4381 accgcggcct tttattgagc ttgaaagata aataaaatag atagggttta tttgaagcta  
4441 aatcttcttt atcgtaaaa atgcctctct ggggttatcaa gagggtcatt atatttcgcg  
4501 gaataacatc atttggtagc gaaataacta agcacttgct tctgtttac tccctgagc  
4561 ttgagggggt aacatgaagg tcatcgatag caggataata atacagtaaa acgctaaacc  
20 4621 aataatccaa atccagccat cccaaattgg tagtgaatga ttataaataa cagcaaacag  
4681 taatgggcca ataaccagg ttgcatgggt aaggctcacc aataatccct gtaaagcacc  
4741 ttgctgatga ctctttgttt ggatagacat cactccctgt aatgcaggta aagcgatccc  
4801 accaccagcc aataaaatta aaacaggga aactaaccaa ccttcagata taaacgctaa  
4861 aaaggcaaat gcactactat ctgcaataaa tccgagcagt actgcggtt tttcgcccat  
4921 ttagtggtta ttcttctgca cacaaggcct tgggaatact agtgtaaaag accaagaccc  
25 4981 gtaatgaaaa gccaacatc atgctattca tcatcagat ttctgtaata gcaccacacc  
5041 gtgctggatt ggctatcaat gcgctgaat aataatcaac aaatggcgc gttaaataag  
5101 tgatgtatag cgtcagctt ttgttccctt tagtgagggt taattgcgcg cttggcgtaa  
5161 tcatgggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata  
30 5221 cgagccggaa gcataaagt taaagcctgg ggtgcctaata gaggtagcta actcacatta  
5281 attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgcgtgcca gctgcattaa  
5341 tgaatcggcc aacgcgcggg gagaggcggt ttgctatttg ggcgtcttc cgttctctcg  
5401 ctactgact cgtcgcgtc ggtcgttcgg ctgcggcgag cggtatcagc tcaactcaag  
5461 gcgtaatac ggttatccac agaatacagg gataacgcag gaaagaacat gtgagcaaaa  
35 5521 ggccagcaaa aggcaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc  
5581 cgccccctg acgagcatca caaaaatcga cgtcaagtc agagggtggcg aaaccgcaca  
5641 ggactataaa gataccaggc gtttccccct ggaagctccc tctgctgctc tctgttccg  
5701 accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct  
5761 catagctcac gctgtaggta tctcagttcg gtgtaggctc ttcgctccaa gctgggctgt  
40 5821 gtgcacgaac ccccggttca gcccgaccgc tgcgcttat ccggtaacta tctgtttag  
5881 tccaaccggg taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc  
5941 agagcgagggt atgtaggcgg tgcacagag ttcttgaagt ggtggcctaa ctacggctac  
6001 actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacott cggaaaaaga  
6061 gttggtagct cttgatccgg aaaaaaacc accgctggtg gcgggtggtt ttttgttgc  
45 6121 aagcagcaga ttacgcgcga aaaaaagga tctcaagaag atcctttgat cttttctacg  
6181 ggggtctgacg ctacgtggaa cgaaaactca cgttaaggga ttttggctat gagattatca  
6241 aaaaggatct tcacctagat ctttttaaat gtttttaaat aatctaaagt  
6301 atatatgagt aaacttggtc tgacagttac caatgcttaa tcaagtaggc acctatctca  
6361 gcgatctgtc tatttcgttc atocatagtt gcctgactcc ccgtcgtgta gataactacg  
50 6421 atacgggagg gcttaccatc tggccccagt gctgcaatga taccgagaga cccacgctca  
6481 ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg cagaagtggg  
6541 cctgcaactt tatccgctc catccagtct attaatgtt gccgggaagc tagagtaagt  
6601 agttcgccag ttaatagttt gcgcaacggt gttgccattg ctacaggcat cgtgggtgca  
6661 cgctcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca  
55 6721 tgatcccca tgtgtgcaa aaaagcgggt agctccttcg gtctccgat cgttgtcaga  
6781 agtaagttgg ccgcagtggt atcactcatg gttatggcag cactgcataa ttctcttact  
6841 gtcattgcat ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga  
6901 gaatagtgtg tgccggcagc gagttgctct tgcccggcgt caatacggga taataccgcg  
6961 ccacatagca gaactttaa agtgctcatc attggaaaac gttcttcggg gcgaaaactc  
60 7021 tcaaggatct taccgctgtt gagatccagt tcatgtaac ccactcgtgc acccaactga  
7081 tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaaat  
7141 gccgcaaaaa aggaataaag ggcgacacgg aaatgttgaa tactcatact cttccttttt  
7201 caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt  
7261 atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccac

65

SEQ ID NO:3 (pTnMod)

	CTGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	50
	CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCC	CTCCTTTCGC	100
	TTTCTTCCCT	TCCTTTCTCG	CCACGTTTCGC	CGGCATCAGA	TTGGCTATTG	150
	GCCATTGCAT	ACGTTGTATC	CATATCATAA	TATGTACATT	TATATTGCCT	200
5	CATGTCCAAC	ATTACCGCCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	250
	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT	AGCCCATATA	TGGAGTTCCG	300
	CGTTACATAA	CTTACGGTAA	ATGGCCCGCC	TGGCTGACCG	CCCAACGACC	350
	CCCGCCCATT	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	400
	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT	AACTGCCCCA	450
10	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTACGCCC	CCTATTGACG	500
	TCAATGACGG	TAAATGGCCC	GCCTGGCATT	ATGCCAGTA	CATGACCTTA	550
	TGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC	600
	CATGGTGATG	CGGTTTTGGC	AGTACATCAA	TGGGCGTGA	TAGCGGTTTG	650
	ACTCACGGGG	ATTTCCAAGT	CTCCACCCCA	TTGACGTCAA	TGGGAGTTTG	700
15	TTTTGGCACC	AAAATCAACG	GGACTTTCCA	AAATGTCGTA	ACAACTCCGC	750
	CCCATTGACG	CAAATGGGCG	GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	800
	GCAGAGCTCG	TTTAGTGAAC	CGTCAGATCG	CCTGGAGACG	CCATCCACGC	850
	TGTTTTGACC	TCCATAGAAG	ACACCGGGAC	CGATCCAGCC	TCCGCGGCCG	900
	GGAACGGTGC	ATTGGAACGC	GGATTCCTCCG	TGCCAAGAGT	GACGTAAGTA	950
20	CCGCCTATAG	ACTCTATAGG	CACACCCCTT	TGGCTCTTAT	GCATGCTATA	1000
	CTGTTTTTGG	CTTGGGGCCT	ATACACCCCC	GCTTCCTTAT	GCTATAGGTG	1050
	ATGGTATAGC	TTAGCCTATA	GGTGTGGGTT	ATTGACCATT	ATTGACCACT	1100
	CCCCTATTGG	TGACGATACT	TTCCATTACT	AATCCATAAC	ATGGCTCTTT	1150
	GCCACAATA	TCTCTATTGG	CTATATGCCA	ATACTCTGTC	CTTCAGAGAC	1200
25	TGACACGGAC	TCTGTATTTT	TACAGGATGG	GGTCCCATTT	ATTATTTACA	1250
	AATTCACATA	TACAACAACG	CCGTCCCCCG	TGCCCGCAGT	TTTTATTAAA	1300
	CATAGCGTGG	GATCTCCACG	CGAATCTCGG	GTACGTGTTC	CGGACATGGG	1350
	CTCTTCTCCG	GTAGCGGCGG	AGCTTCCACA	TCCGAGCCCT	GGTCCCATGC	1400
	CTCCAGCGGC	TCATGGTTCG	TCCGCGAGCTC	CTTGCTCCTA	ACAGTGGAGG	1450
30	CCAGACTTAG	GCACAGCACA	ATGCCACCA	CCACCAGTGT	GCCGCACAAG	1500
	GCCGTGGCGG	TAGGGTATGT	GTCTGAAAAT	GAGCGTGGAG	ATTGGGCTCG	1550
	CACGGCTGAC	GCAGATGGAA	GACTTAAGGC	AGCGGCAGAA	GAAGATGCAG	1600
	GCAGCTGAGT	TGTTGTATTC	TGATAAGAGT	CAGAGGTAAC	TCCCGTTGCG	1650
	GTGCTGTTAA	CGGTGGAGGG	CAGTGTAGTC	TGAGCAGTAC	TCGTTGCTGC	1700
35	CGCGCGCGCC	ACCAGACATA	ATAGCTGACA	GACTAACAGA	CTGTTCCCTT	1750
	CCATGGGTCT	TTTCTGCAGT	CACCGTCGGA	CCATGTGTGA	ACTTGATATT	1800
	TTACATGATT	CTCTTTACCA	ATTCTGCCCC	GAATTACACT	TAAAACGACT	1850
	CAACAGCTTA	ACGTTGGCTT	GCCACGCATT	ACTTGACTGT	AAAACCTCTA	1900
	CTCTTACCGA	ACTTGGCCGT	AACCTGCCAA	CCAAAGCGAG	AACAAAACAT	1950
40	AACATCAAAC	GAATCGACCG	ATTGTTAGGT	AATCGTCACC	TCCACAAAGA	2000
	GCGACTCGCT	GTATACCGTT	GGCATGCTAG	CTTTATCTGT	TCGGGAATAC	2050
	GATGCCCATT	GTACTTGTTG	ACTGGTCTGA	TATTCTGTGAG	CAAAAACGAC	2100
	TTATGGTATT	GCGAGCTTCA	GTCGCACTAC	ACGGTCGTTC	TGTTACTCTT	2150
	TATGAGAAAG	CGTTCCCGCT	TTCAGAGCAA	TGTTCAAAGA	AAGCTCATGA	2200
45	CCAATTTCTA	GCCGACCTTG	CGAGCATTCT	ACCGAGTAAC	ACCACACCGC	2250
	TCATTGTCAG	TGATGCTGGC	TTTAAAGTGC	CATGGTATAA	ATCCGTTGAG	2300
	AAGCTGGGTT	GGTACTGGTT	AAGTCGAGTA	AGAGGAAAAG	TACAATATGC	2350
	AGACCTAGGA	GCGGAAAAC	GGAAACCTAT	CAGCAACTTA	CATGATATGT	2400
	CATCTAGTCA	CTCAAAGACT	TTAGGCTATA	AGAGGCTGAC	TAAAAGCAAT	2450
50	CCAATCTCAT	GCCAAATTCT	ATTGTATAAA	TCTCGCTCTA	AAGGCCGAAA	2500
	AAATCAGCGC	TCGACACGGA	CTCATTGTCA	CCACCCGTCA	CCTAAAATCT	2550
	ACTCAGCGTC	GGCAAAGGAG	CCATGGGTTC	TAGCAACTAA	CTTACCTGTT	2600
	GAAATTCGAA	CACCCAAACA	ACTTGTTAAT	ATCTATTGCA	AGCGAATGCA	2650
	GATTGAAGAA	ACCTTCCGAG	ACTTGAAAAG	TCCTGCCTAC	GGACTAGGCC	2700
55	TACGCCATAG	CCGAACGAGC	AGCTCAGAGC	GTTTTGATAT	CATGCTGCTA	2750
	ATCGCCCTGA	TGCTTCAACT	AACATGTTGG	CTTGCGGGCG	TTCATGCTCA	2800
	GAAACAAGGT	TGGGACAAGC	ACTTCCAGGC	TAACACAGTC	AGAAATCGAA	2850
	ACGTACTCTC	AACAGTTCGC	TTAGGCATGG	AAGTTTTGCG	GCATTCTGGC	2900
	TACACAATAA	CAAGGGAAGA	CTTACTCGTG	GCTGCAACCC	TACTAGCTCA	2950
60	AAATTTATTCT	ACACATGGTT	ACGCTTTGGG	GAAATTATGA	TAATGATCCA	3000
	GATCACTTCT	GGCTAATAAA	AGATCAGAGC	TCTAGAGATC	TGTGTGTTGG	3050



TTTTTTGTGG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC TGTGTTTGC 3100  
CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC CCACTGTCCT 3150  
TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT 3200  
CTATTCTGGG GGGTGGGGTG GGGCAGCACA GCAAGGGGGA GGATTGGGAA 3250  
5 GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCTCTCT 3300  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCGGTAC CTCTCTCTCT 3350  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CGGTACCAGG TGCTGAAGAA 3400  
TTGACCCGGT GACCAAAGGT GCCTTTTATC ATCACTTTAA AAATAAAAAA 3450  
CAATTACTCA GTGCCTGTTA TAAGCAGCAA TTAATTATGA TTGATGCCTA 3500  
10 CATCACAACA AAAACTGATT TAACAAATGG TTGGTCTGCC TTAGAAAGTA 3550  
TATTTGAACA TTATCTTGAT TATATTATTG ATAATAATAA AAACCTTATC 3600  
CCTATCCAAG AAGTGATGCC TATCATTGGT TGGAATGAAC TTGAAAAAAA 3650  
TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATTG 3700  
ATCCAAGAGA ACCAACTTAA AGCTTTCCTG ACGGAATGTT AATTCTCGTT 3750  
15 GACCCTGAGC ACTGATGAAT CCCCTAATGA TTTTGGTAAA AATCATTAAG 3800  
TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850  
CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT 3900  
TGTGTGGAAT TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC 3950  
CATGATTACG CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT 4000  
20 GGAGCTCCAC CGCGGTGGCG GCCGCTCTAG AACTAGTGA TCCCCGGGGC 4050  
TGCAGGAATT CGATATCAAG CTTATCGATA CCGCTGACCT CGAGGGGGGG 4100  
CCCGGTACCC AATTCGCCCT ATAGTGAGTC GTATTACGCG CGCTCACTGG 4150  
CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACCT 4200  
AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA 4250  
25 GGCCCGCACC GATCGCCCTT CCAACAGTT GCGCAGCCTG AATGGCGAAT 4300  
GGAAATTGTA AGCGTTAATA TTTTGTAAA ATTGCGGTTA AATTTTGT 4350  
AAATCAGCTC ATTTTTTAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT 4400  
AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTGTTT CAGTTTGGAA 4450  
CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAA GGGCGAAAAA 4500  
30 CCGTCTATCA GGGCGATGGC CCACTACTCC GGGATCATAT GACAAGATGT 4550  
GTATCCACCT TAACCTAATG ATTTTACCA AAATCATTAG GGGATTATC 4600  
AGTGCTCAGG GTCAACGAGA ATTAACATTC CGTCAGGAAA GCTTATGATG 4650  
ATGATGTGCT TAAAACTTA CTCAATGGCT GGTTATGCAT ATCGCAATAC 4700  
ATGCGAAAAA CCTAAAAGAG CTTGCCGATA AAAAAGGCCA ATTTATTGCT 4750  
35 ATTTACGCGG GCTTTTATT GAGCTTGAAA GATAAATAAA ATAGATAGGT 4800  
TTTATTTGAA GCTAAATCTT CTTTATCGTA AAAAATGCCC TCTTGGGTTA 4850  
TCAAGAGGGT CATTATATTT CGCGGAATAA CATCATTGGG TGACGAAATA 4900  
ACTAAGCACT TGTCTCCTGT TTACTCCCTT GAGCTTGAGG GGTAAACATG 4950  
AAGGTCATCG ATAGCAGGAT AATAATACAG TAAAACGCTA AACCAATAAT 5000  
40 CCAAATCCAG CCATCCCAA TTTGGTAGTA ATGATTATAA ATAACAGCAA 5050  
ACAGTAATGG GCCAATAACA CCGGTGTCAT TGGTAAGGCT CACCAATAAT 5100  
CCCTGTAAAG CACCTTGCTG ATGACTCTTT GTTTGGATAG ACATCACTCC 5150  
CTGTAATGCA GGTAAAGCGA TCCCACCACC AGCCAATAAA ATTAAACAG 5200  
GGAAACTTAA CCAACCTTCA GATATAACG CTAAAAGGC AAATGCACTA 5250  
45 CTATCTGCAA TAAATCCGAG CAGTACTGCC GTTTTTTCGC CCATTTAGTG 5300  
GCTATTCTTC CTGCCACAAA GGCTTGGAAT ACTGAGTGTA AAAGACCAAG 5350  
ACCCGTAATG AAAAGCCAAC CATCATGCTA TTCATCATCA CGATTTCTGT 5400  
AATAGCACCA CACCGTGCTG GATTGGCTAT CAATGCGCTG AAATAATAAT 5450  
CAACAAATGG CATCGTTAAA TAAGTGATGT ATACCGATCA GCTTTTGTTC 5500  
50 CCTTTAGTGA GGGTTAATTG CGCGCTTGGC GTAATCATGG TCATAGCTGT 5550  
TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC 5600  
GGAAGCATAA AGTGTAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC 5650  
ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT 5700  
GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCCT 5750  
55 ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT 5800  
TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT 5850  
CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG 5900  
CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG 5950  
GCTCCGCCCC CCTGACGAGC ATCACAATAA TCGACGCTCA AGTCAGAGGT 6000  
60 GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC 6050  
TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC 6100

5 CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA 6150  
 GGTATCTCAG TTCGGTGTAG GTCGTTGCTT CCAAGCTGGG CTGTGTGCAC 6200  
 GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT 6250  
 TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG 6300  
 10 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG 6350  
 AAGTGGTGGC CTAACACGG CTACACTAGA AGGACAGTAT TTGGTATCTG 6400  
 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT 6450  
 CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG 6500  
 CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC 6550  
 15 TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTAA GGGATTTTGG 6600  
 TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA 6650  
 TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAACTT GGTCTGACAG 6700  
 TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTC 6750  
 GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG 6800  
 20 GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG 6850  
 CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG 6900  
 AGCGCAGAAG TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT 6950  
 TGTGCGCGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA 7000  
 CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA 7050  
 25 TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC 7100  
 CCCATGTTGT GCAAAAAGC GGTAGCTCC TTCGGTCTC CGATCGTTGT 7150  
 CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC 7200  
 ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT 7250  
 GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG 7300  
 30 CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT 7350  
 TAAAGTGCT CATCATTGGA AAACGTTCTT CCGGGCGAAA ACTCTCAAGG 7400  
 ATCTTACCGC TGTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA 7450  
 CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA 7500  
 CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT 7550  
 35 TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG 7600  
 TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC 7650  
 AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCAC 7689

35 SEQ ID NO:4 (a Kozak sequence)  
 ACCATGG

SEQ ID NO:5 (a Kozak sequence)  
 ACCATGT

40 SEQ ID NO:6 (a Kozak sequence)  
 AAGATGT

45 SEQ ID NO:7 (a Kozak sequence)  
 ACGATGA

SEQ ID NO:8 (a Kozak sequence)  
 AAGATGG

50 SEQ ID NO:9 (a Kozak sequence)  
 GACATGA

SEQ ID NO:10 (a Kozak sequence)  
 ACCATGA

55 SEQ ID NO:11 (conalbumin polyA)  
 tctgccattg ctgcttcctc tgcccttcct cgctactctg aatgtggctt ctctgctact  
 gccacagcaa gaaataaaat ctcaacatct aaatggggtt cctgaggttt ttcaagagtc

5

10

15

20

25

30

35

40

45

50

55

60

GGGAGAGTAT AGTCAATGTA  
 GATTATACTG ATTGCTGATT  
 GGGTTATTAT CAGCTAGATA  
 ACAACTTGGG TCAGGTGCCA  
 5 GGTCAACATA ACCTGGGCAA  
 AACCAGTCTC ATCTGTGGCA  
 GGACCATGTA CCAGCAGCCA  
 GCCGTGACCC AATCTAGGAA  
 AGCAAGTAGC ACATCAATTT  
 10 TAAATTTATT GTAAATGCCG  
 TAGTAGAAGT GTTTTACTGT  
 GATACATTGA AACTTCTGGT  
 CAATCAGAAA AAGGTTTTTT  
 ATCAGAGATG CCAAGGTATT  
 15 ATTTGATTTT CTTTATTCGC  
 CGTGAAGAGA ATTTATGATT  
 GCAAAAAGAG GAGTGTTTAC  
 ATAAACTGAT AAAAACTTG  
 AGGAATTCAG CAGAAAACAG  
 20 CCACGTGTTT CTGAACATTC  
 TTCCATAAAA GTCTCACCAT  
 GCCTGGCAGA GCCCTATTCA  
 CCTTCGCT

25 SEQ ID NO:15 (fragment of ovalbumin promoter - chicken)  
 GAGGTCAGAAT GGTTTCTTTA CTGTTTGTC AATCTATTAT TTCAATACAG  
 AACAAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG  
 TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTACAA TTCCTCTGTC  
 30 ATCTGCCAGG CCATTAAGTT ATTCATGGAA GATCTTTGAG GAACACTGCA  
 AGTTCATATC ATAAACACAT TTGAAATTGA GTATTGTTTT GCATTGTATG  
 GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTTA TAAAGCATTG  
 ACACCCATAA AAAGATAGAT TTAAATATTC CAGCTATAGG AAAGAAAGTG  
 CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCTTCACAT GCATGCTTCT  
 35 TTATTTCTCC TATTTTGTC AAAAAATAAT AGGTCACGTC TTGTCTCAC  
 TTATGTCCTG CCTAGCATGG CTCAGATGCA CGTTGTAGAT ACAAGAAGGA  
 TCAAATGAAA CAGACTTCTG GTCTGTTACT ACAACCATAG TAATAAGCAC  
 ACTAATAAT AATTGCTAAT TATGTTTTCC ATCTCTAAGG TTCCACATT  
 TTTCTGTTTT CTAAAGATC CCATTATCTG GTTGTAAGT AAGCTCAATG  
 40 GAACATGAGC AATATTTCCC AGTCTTCTCT CCCATCCAAC AGTCCTGATG  
 GATTAGCAGA ACAGGCAGAA AACACATTGT TACCCAGAAT TAAAACTAA  
 TATTTGCTCT CCATTCAATC CAAAATGGAC CTATTGAAAC TAAAATCTAA  
 CCAATCCCA TTAAATGATT TCTATGGCGT CAAAGGTCAA ACTTCTGAAG  
 45 GGAACCTGTG GGTGGGTCAC AATTCAGGCT ATATATTTCC CAGGGCTCAG

SEQ ID NO:16 (chicken ovalbumin enhancer)  
 ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag  
 cttgacctga tacctgattt tcttcaaact ggggaaacaa cacaatccca caaaacagct  
 50 cagagagaaa ccatcactga tggctacagc accaaggat gcaatggcaa tccattcgac  
 attcatctgt gacctgagca aaatgattta tctctccatg aatgggtgct tcttccctc  
 atgaaaaggc aatttccaca ctcaaatat gcaacaaaga caaacagaga acaattaatg  
 tgctccttcc taatgtcaaa attgtagtgg caaagaggag acaaaaatct caagttctga  
 gtaggtttta gtgattggat aagaggcttt gacctgtgag ctacctgga cttcatatcc  
 55 ttttggataa aaagtgttt tataactttc aggtctccga gtctttatc atgagactgt  
 tggtttaggg acagaccac aatgaaatgc ctggcatagg aaagggcagc agagccttag  
 ctgacctttt cttgggacaa gcattgtcaa acaatgtgtg acaaaaactat ttgtactgct  
 ttgcacagct gtgctgggca gggcaatcca ttgccaccta tcccaggtaa cttccaact  
 60 gcaagaagat tggtgcttac tctctctaga

SEQ ID NO:17 (5' untranslated region)

GTGGATCAACATACAGCTAGAAAGCTGTATTGCCTTTAGCACTCAAGCTCAAAGACAACCTCAGAGTTC  
ACC

5

SEQ ID NO:18 (putative cap site)

ACATACAGCTAG AAAGCTGTAT TGCCTTTAGC ACTCAAGCTC AAAAGACAAC TCAGAGTTCA

10

SEQ ID NO:19 (Chicken Ovalbumin Signal Sequence)

ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA GCTCAAAGTC  
CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TCATGTCAGC TCTAGCCATG  
GTATACCTGG GTGCAAAAGA CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT  
AAACTTCCAG GATTCCGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTTAC  
15 TCTTCACTTA GAGACATCCT CAACCAAATC ACCAAACCAA ATGATGTTTA TTCGTTTACG  
CTTGCCAGTA GACTTTATGC TGAAGAGAGA TACCCAATCC TGCCAGAATA CTTGCAGTGT  
GTGAAGGAAC TGTATAGAGG AGGCTTGGA CCTATCAACT TTCAAACAGC TGCAGATCAA  
GCCAGAGAGC TCATCAATTC CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC  
CTTCAGCCAA GCTCCGTGGA TTCTCAAAC GCAATGGTTC TGGTTAATGC CATTGTCTTC  
20 AAAGGACTGT GGGAGAAAAC ATTTAAGGAT GAAGACACAC AAGCAATGCC TTTCAGAGTG  
ACTGAGCAAG AAAGCAAACC TGTGCAGATG ATGTACCAGA TTGGTTTATT TAGAGTGGCA  
TCAATGGCTT CTGAGAAAAT GAAGATCCTG GAGCTTCCAT TTGCCAGTGG GACAATGAGC  
ATGTTGGTGC TGTTCCTGA TGAAGTCTCA GGCCTTGAGC AGCTTGAGAG TATAATCAAC  
TTTGAAAAAC TGACTGAATG GACCAGTTCT AATGTTATGG AAGAGAGGAA GATCAAAGTG  
25 TACTTACCTC GCATGAAGAT GGAGGAAAAA TACAACCTCA CATCTGTCTT AATGGCTATG  
GGCATTACTG ACGTGTTTAG CTCTTCAGCC AATCTGTCTG GCATCTCCTC AGCAGAGAGC  
CTGAAGATAT CTAAGCTGT CCATGCAGCA CATGCAGAAA TCAATGAAGC AGGCAGAGAG  
GTGGTAGGGT CAGCAGAGGC TGGAGTGGAT GCTGCAAGCG TCTCTGAAGA ATTTAGGGCT  
GACCATCCAT TCCTCTTCTG TATCAAGCAC ATCGCAACCA ACGCCGTTCT CTCTTTTGGC  
30 AGATGTGTTT CCCCT

SEQ ID NO:20 (Chicken Ovalbumin Signal Sequence - shortened 50bp)

ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA

35

SEQ ID NO:21 (Chicken Ovalbumin Signal Sequence - shortened 100bp)

ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTTGATG TATTCAAGGA GCTCAAAGTC  
CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA

40

SEQ ID NO:22 (vitellogenin targeting sequence)

ATGAGGGGGATCATACTGGCATTAGTGCTCACCCCTGTAGGCAGCCAGAAGTTTGACATTGGT

45

SEQ ID NO:23 (pro-insulin sequence)

TTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGGAAGCTCTCTACCTAGTGTGCGGGGAACGAGGC  
TTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGG  
GGCCCTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAA  
50 CAATGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAAGTCTGCAACTAG

SEQ ID NO:24 (p146 protein)

KYKKALKKLAKLL

55

SEQ ID NO:25 (p146 coding sequence)

AAATACAAAAAGCACTGAAAAAAGTGGCAAACTGCTG

60

SEQ ID NO:26 (spacer)



(GPGG)<sub>x</sub>

SEQ ID NO:27 (spacer)  
GPGGGPGGGPGG

5

SEQ ID NO:28 (spacer)  
GGGSGGGSGGGGS

10

SEQ ID NO:29 (spacer)  
GGGSGGGSGGGSGGGGS

15 SEQ ID NO:30 (repeat domain in TAG spacer sequence)  
Pro Ala Asp Asp Ala

20 SEQ ID NO:31 (TAG spacer sequence)  
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp  
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp

25 SEQ ID NO:32 (gp41 epitope)  
Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu

30 SEQ ID NO:33 (polynucleotide sequence encoding gp41 epitope)  
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly  
Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys

35 SEQ ID NO:34 (enterokinase cleavage site)  
DDDDK

40 SEQ ID NO:35 (TAG sequence)  
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp  
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys  
Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys

45 SEQ ID NO:36 (altered transposase Hef forward primer)  
ATCTCGAGACCATGTGTGAACTTGATATTTTACATGATTCTCTTTACC

50 SEQ ID NO:37 (altered transposase Her reverse primer)  
GATTGATCATTATCATAATTTCCCAAAGCGTAACC

55 SEQ ID NO:38 (Xho I restriction site)  
CTCGAG

SEQ ID NO:39 (Bcl I restriction site)  
TGATCA

SEQ ID NO:40 (CMVf-NgoM IV primer)  
TTGCCGGCATCAGATTGGCTAT

5

SEQ ID NO:41 (Syn-polyAr-BstE II primer)  
AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA

10 SEQ ID NO:42 (pTnMod(Oval/ENT tag/Proins/PA) - Chicken)  
CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50  
CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTTCGC 100  
TTTCTTCCCT TCCTTTCTCG CCACGTTCGC CGGCATCAGA TTGGCTATTG 150  
GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200  
15 CATGTCCAAC ATTACCGCCA TGTTGACATT GATTATTGAC TAGTTATTAA 250  
TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300  
CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC 350  
CCCGCCCAT TACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400  
GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAAC TGCCCA 450  
20 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC CCTATTGACG 500  
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 550  
TGGGACTTTC CTAATTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600  
CATGGTGATG CGGTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG 650  
ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700  
25 TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC 750  
CCCATTTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 800  
GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850  
TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900  
GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA 950  
30 CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000  
CTGTTTTTGG CTTGGGGCCT ATACACCCCC GCTTCTTAT GCTATAGGTG 1050  
ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCATT ATTGACCACT 1100  
CCCCATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150  
GCCACAATA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTGAGAGAC 1200  
35 TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATTATTTACA 1250  
AATTCACATA TACAACAACG CCGTCCCCCG TGCCCGCAGT TTTTATTAAA 1300  
CATAGCGTGG GATCTCCACG CGAATCTCGG GTACGTGTTT CCGACATGGG 1350  
CTCTTCTCCG GTAGCGGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400  
CTCCAGCGGC TCATGGTCGC TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450  
40 CCAGACTTAG GCACAGCACA ATGCCACCA CCACCAGTGT GCCGCACAAG 1500  
GCCGTGGCGG TAGGGTATGT GTCTGAAAAT GAGCGTGGAG ATTGGGCTCG 1550  
CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600  
GCAGCTGAGT TGTGTATTC TGATAAGAGT CAGAGGTAAC TCCCGTTGCG 1650  
GTGCTGTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700  
45 CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCTTTT 1750  
CCATGGGTCT TTTCTGCAGT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800  
TTACATGATT CTCTTTACCA ATTCTGCCCC GAATTACACT TAAAACGACT 1850  
CAACAGCTTA ACGTTGGCTT GCCACGCATT ACTTGACTGT AAAACTCTCA 1900  
CTCTTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCGAG AACAAAACAT 1950  
50 AACATCAAAC GAATCGACCG ATTGTTAGGT AATCGTCACC TCCACAAAGA 2000  
GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050  
GATGCCCAT GTACTTGTG ACTGGTCTGA TATTCGTGAG CAAAAACGAC 2100  
TTATGGTATT GCGAGCTTCA GTCGCACTAC ACGGTCGTTC TGTTACTCTT 2150  
TATGAGAAAG CGTTCCTCGT TTCAGAGCAA TGTTCAAAGA AAGCTCATGA 2200  
55 CCAATTTCTA GCCGACCTTG CGAGCATTCT ACCGAGTAAC ACCACACCGC 2250  
TCATTGTCAG TGATGCTGGC TTTAAAGTGC CATGGTATAA ATCCGTTGAG 2300  
AAGCTGGGTT GGTACTGGTT AAGTCGAGTA AGAGGAAAAG TACAATATGC 2350  
AGACCTAGGA GCGGAAACT GGAAACCTAT CAGCAACTTA CATGATATGT 2400  
CATCTAGTCA CTCAAAGACT TTAGGCTATA AGAGGCTGAC TAAAAGCAAT 2450  
60 CCAATCTCAT GCCAAATTCT ATTGTATAAA TCTCGCTCTA AAGGCCGAAA 2500  
AAATCAGCGC TCGACACGGA CTCATTGTCA CCACCCGTCA CCTAAAATCT 2550

ACTCAGCGTC GGCAAAGGAG CCATGGGTTT TAGCAACTAA CTTACCTGTT 2600  
GAAATTCGAA CACCCAAACA ACTTGTTAAT ATCTATTCGA AGCGAATGCA 2650  
GATTGAAGAA ACCTTCCGAG ACTTGAAAAG TCCTGCCTAC GGACTAGGCC 2700  
TACGCCATAG CCGAACGAGC AGCTCAGAGC GTTTTGATAT CATGCTGCTA 2750  
5 ATCGCCCTGA TGCTTCAACT AACATGTTGG CTTGCGGGCG TTCATGCTCA 2800  
GAAACAAGGT TGGGACAAGC ACTTCCAGGC TAACACAGTC AGAAATCGAA 2850  
ACGTACTCTC AACAGTTCGC TTAGGCATGG AAGTTTTCGC GCATTCTGGC 2900  
TACACAATAA CAAGGGAAGA CTTACTCGTG GCTGCAACCC TACTAGCTCA 2950  
AAATTTATTC ACACATGGTT ACGCTTTGGG GAAATTATGA TAATGATCCA 3000  
10 GATCACTTCT GGCTAATAAA AGATCAGAGC TCTAGAGATC TGTGTGTTGG 3050  
TTTTTTGTGG ATCTGCTGTG CTTTCTAGTT GCCAGCCATC TGTGTGTTGC 3100  
CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC CCCTGTCTCT 3150  
TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT 3200  
CTATTCTGGG GGGTGGGGTG GGGCAGCACA GCAAGGGGGA GGATTGGGAA 3250  
15 GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCTCTCT 3300  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCGGTAC CTCTCTCTCT 3350  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CCGTACCAGG TGCTGAAGAA 3400  
TTGACCCGGT GACCAAAGGT GCCTTTTATC ATCACTTTAA AAATAAAAAA 3450  
CAATTACTCA GTGCCTGTTA TAAGCAGCAA TTAATTATGA TTGATGCCTA 3500  
20 CATCACAACA AAAACTGATT TAACAAATGG TTGGTCTGCC TTAGAAAGTA 3550  
TATTTGAACA TTATCTTGAT TATATTATTG ATAATAATAA AAACCTTATC 3600  
CCTATCCAAG AAGTGATGCC TATCATTGGT TGAATGAAC TTGAAAAAAA 3650  
TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATTG 3700  
ATCCAAGAGA ACCAACTTAA AGCTTTCCTG ACGGAATGTT AATTCTCGTT 3750  
25 GACCCTGAGC ACTGATGAAT CCCCTAATGA TTTTGGTAAA AATCATTAAAG 3800  
TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850  
CACTCATTAG GCACCCGAGG CTTTACACTT TATGCTTCCG GCTCGTATGT 3900  
TGTGTGGAAT TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC 3950  
CATGATTACG CCAAGCGCGC AATTAACCTT CACTAAAGGG AACAAAAGCT 4000  
30 GGAGCTCCAC CGCGGTGGCG GCCGCTCTAG AACTAGTGGA TCCCCGGGG 4050  
AGGTCAGAAT GGTTTCTTTA CTGTTTGTCA ATTCTATTAT TTCAATACAG 4100  
ACAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG 4150  
TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTACAA TTCCTCTGTC 4200  
ATCTGCCAGG CCATTAAGTT ATTCATGGAA GATCTTTGAG GAACACTGCA 4250  
35 AGTTCATATC ATAAACACAT TTGAAATTGA GTATTGTTTT GCATTGTATG 4300  
GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTTA TAAAGCATTC 4350  
ACACCCATAA AAAGATAGAT TTAAATATTC CAGCTATAGG AAAGAAAGTG 4400  
CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCFTCACAT GCATGCTTCT 4450  
TTATTTCTCC TATTTTGTCA AGAAAATAAT AGGTCACGTC TTGTTCTCAC 4500  
40 TTATGTCCTG CCTAGCATGG CTCAGATGCA CGTTGTAGAT ACAAGAAGGA 4550  
TCAAATGAAA CAGACTTCTG GTCTGTTACT ACAACCATAG TAATAAGCAC 4600  
ACTAACTAAT AATTGCTAAT TATGTTTTCC ATCTCTAAGG TTCCACATT 4650  
TTTCTGTTTT CTAAAGATC CCATTATCTG GTTGTAAGT AAGCTCAATG 4700  
GAACATGAGC AATATTCCC AGTCTTCTCT CCCATCCAAC AGTCCTGATG 4750  
45 GATTAGCAGA ACAGGCAGAA AACACATTGT TACCCAGAAT TAAAACTAA 4800  
TATTTGCTCT CCATTCAATC CAAAATGGAC CTATTGAAAC TAAATCTAA 4850  
CCCAATCCCA TTAAATGATT TCTATGGCGT CAAAGGTCAA ACTTCTGAAG 4900  
GGAACCTGTG GGTGGGTCAC AATTCAGGCT ATATATTCCC CAGGGCTCAG 4950  
CGGATCCATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTGTTGATG 5000  
50 TATTCAAGGA GCTCAAAGTC CACCATGCCA ATGAGAACAT CTTCTACTGC 5050  
CCCATTGCCA TCATGTCAGC TCTAGCCATG GTATACCTGG GTGCAAAAGA 5100  
CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT AAACCTCCAG 5150  
GATTCCGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTTAC 5200  
TCTTCACTTA GAGACATCCT CAACCAAATC ACCAAACCAA ATGATGTTTA 5250  
55 TTCGTTGAGC CTTGCCAGTA GACTTTATGC TGAAGAGAGA TACCCAATCC 5300  
TGCCAGAATA CTTGCAGTGT GTGAAGGAAC TGTATAGAGG AGGCTTGGA 5350  
CCTATCAACT TTCAAACAGC TGCAGATCAA GCCAGAGAGC TCATCAATTC 5400  
CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC CTTAGCCAA 5450  
GCTCCGTGGA TTCTCAAAC GCAATGGTTC TGGTTAATGC CATTGTCTTC 5500  
60 AAAGGACTGT GGGAGAAAAC ATTTAAGGAT GAAGACACAC AAGCAATGCC 5550  
TTTCAGAGTG ACTGAGCAAG AAAGCAAACC TGTGCAGATG ATGTACCAGA 5600

	TTGGTTTATT	TAGAGTGGCA	TCAATGGCTT	CTGAGAAAAT	GAAGATCCTG	5650
	GAGCTTCCAT	TTGCCAGTGG	GACAATGAGC	ATGTTGGTGC	TGTTGCCTGA	5700
	TGAAGTCTCA	GGCCTTGAGC	AGCTTGAGAG	TATAATCAAC	TTTGAAAAAC	5750
	TGACTGAATG	GACCAGTTCT	AATGTTATGG	AAGAGAGGAA	GATCAAAGTG	5800
5	TACTTACCTC	GCATGAAGAT	GGAGGAAAAA	TACAACCTCA	CATCTGTCTT	5850
	AATGGCTATG	GGCATTACTG	ACGTGTTTAG	CTCTTCAGCC	AATCTGTCTG	5900
	GCATCTCCTC	AGCAGAGAGC	CTGAAGATAT	CTCAAGCTGT	CCATGCAGCA	5950
	CATGCAGAAA	TCAATGAAGC	AGGCAGAGAG	GTGGTAGGGT	CAGCAGAGGC	6000
	TGGAGTGGAT	GCTGCAAGCG	TCTCTGAAGA	ATTTAGGGCT	GACCATCCAT	6050
10	TCCTCTTCTG	TATCAAGCAC	ATCGCAACCA	ACGCCGTTCT	CTTCTTTGGC	6100
	AGATGTGTTT	CCCCTCCGCG	GCCAGCAGAT	GACGCACCAG	CAGATGACGC	6150
	ACCAGCAGAT	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	GACGCACCAG	6200
	CAGATGACGC	AACAACATGT	ATCCTGAAAG	GCTCTTGTTG	CTGGATCGGC	6250
	CTGCTGGATG	ACGATGACAA	ATTTGTGAAC	CAACACCTGT	GCGGCTCACA	6300
15	CCTGGTGGAA	GCTCTCTACC	TAGTGTGCGG	GGAACGAGGC	TTCTTCTACA	6350
	CACCCAAGAC	CCGCCGGGAG	GCAGAGGACC	TGCAGGTGGG	GCAGGTGGAG	6400
	CTGGGCGGGG	GCCCTGGTGC	AGGCAGCCTG	CAGCCCTTGG	CCCTGGAGGG	6450
	GTCCCTGCAG	AAGCGTGGCA	TTGTGGAACA	ATGCTGTACC	AGCATCTGCT	6500
	CCCTCTACCA	GCTGGAGAAC	TACTGCAACT	AGGGCGCCTG	GATCCAGATC	6550
20	ACTTCTGGCT	AATAAAAGAT	CAGAGCTCTA	GAGATCTGTG	TGTTGGTTTT	6600
	TTGTGGATCT	GCTGTGCCCT	CTAGTTGCCA	GCCATCTGTT	GTTTGCCCCT	6650
	CCCCCGTGCC	TTCTTTGACC	CTGGAAGGTG	CCACTCCCAC	TGTCCTTTCC	6700
	TAATAAAATG	AGGAAATTGC	ATCGCATTGT	CTGAGTAGGT	GTCATTCTAT	6750
	TCTGGGGGGT	GGGGTGGGGC	AGCACAGCAA	GGGGGAGGAT	TGGGAAGACA	6800
25	ATAGCAGGCA	TGCTGGGGAT	GCGGTGGGCT	CTATGGGTAC	CTCTCTCTCT	6850
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCTCT	CTCGAGGGGG	6900
	GGCCCGGTAC	CCAATTCGCC	CTATAGTGAG	TCGTATTACG	CGCGCTCACT	6950
	GGCCGTCTG	TTACAACGTC	GTGACTGGGA	AAACCCTGGC	GTTACCCAAC	7000
	TTAATCGCCT	TGCAGCACAT	CCCCCTTTCC	CCAGCTGGCG	TAATAGCGAA	7050
30	GAGGCCCGCA	CCGATCGCCC	TTCCCAACAG	TTGCGCAGCC	TGAATGGCGA	7100
	ATGGAAATTG	TAAGCGTTAA	TATTTTGTTA	AAATTGCGGT	TAAATTTTTG	7150
	TTAAATCAGC	TCATTTTTTA	ACCAATAGGC	CGAAATCGGC	AAAATCCCTT	7200
	ATAAATCAAA	AGAATAGACC	GAGATAGGGT	TGAGTGTGTT	TCCAGTTTGG	7250
	AACAAGAGTC	CACTATTAAA	GAACGTGGAC	TCCAACGTCA	AAGGGCGAAA	7300
35	AAACGTCTAT	CAGGGCGATG	GCCCACTACT	CCGGGATCAT	ATGACAAGAT	7350
	GTGTATCCAC	CTTAACCTAA	TGATTTTTAC	CAAAATCATT	AGGGGATTCA	7400
	TCAGTGCTCA	GGGTCAACGA	GAATTAACAT	TCCGTCAGGA	AAGCTTATGA	7450
	TGATGATGTG	CTTAAAAACT	TACTCAATGG	CTGGTTATGC	ATATCGCAAT	7500
	ACATGCGAAA	AACCTAAAAG	AGCTTGCCGA	TAAAAAAGGC	CAATTTATTG	7550
40	CTATTTACCG	CGGCTTTTTA	TTGAGCTTGA	AAGATAAATA	AAATAGATAG	7600
	GTTTTATTGG	AAGCTAAATC	TTCTTTATCG	TAAAAAATGC	CCTCTTGGGT	7650
	TATCAAGAGG	GTCATTATAT	TTGCGGGAAT	AACATCATTT	GGTGACGAAA	7700
	TAACCTAAGCA	CTTGTCTCCT	GTTTACTCCC	CTGAGCTTGA	GGGGTTAACA	7750
	TGAAGGTCAT	CGATAGCAGG	ATAATAATAC	AGTAAAACGC	TAAACCAATA	7800
45	ATCCAAATCC	AGCCATCCCA	AATGGTGTAGT	GAATGATTAT	AAATAACAGC	7850
	AAACAGTAAT	GGGCCAATAA	CACCGGTTGC	ATTGGTAAGG	CTCACCAATA	7900
	ATCCCTGTAA	AGCACCTTGC	TGATGACTCT	TTGTTTGGAT	AGACATCACT	7950
	CCCTGTAAATG	CAGGTAAAGC	GATCCCACCA	CCAGCCAATA	AAATTAAAAC	8000
	AGGGAAAACT	AACCAACCTT	CAGATATAAA	CGCTAAAAAG	GCAAATGCAC	8050
50	TACTATCTGC	AATAAATCCG	AGCAGTACTG	CCGTTTTTTC	GCCCCATTTA	8100
	GTGGCTATTG	TTCTTGCCAC	AAAGGCTTGG	AATACTGAGT	GTAAAAGACC	8150
	AAGACCCGCT	AATGAAAAGC	CAACCATCAT	GCTATTCCAT	CCAAAACGAT	8200
	TTTCGGTAAA	TAGCACCAC	ACCGTTGCGG	GAATTTGGCC	TATCAATTGC	8250
	GCTGAAAAAT	AAATAATCAA	CAAAATGGCA	TCGTTTTAAA	TAAAGTGATG	8300
55	TATACCGAAT	TCAGCTTTTG	TTCCCTTTAG	TGAGGGTTAA	TTGCGCGCTT	8350
	GGCGTAATCA	TGGTCATAGC	TGTTTCCTGT	GTGAAATTGT	TATCCGCTCA	8400
	CAATTCCACA	CAACATACGA	GCCGGAAGCA	TAAAGTGTA	AGCCTGGGGT	8450
	GCCTAATGAG	TGAGCTAACT	CACATTAATT	GCGTTGCGCT	CACTGCCCCG	8500
	TTTCCAGTCG	GGAAACCTGT	CGTGCCAGCT	GCATTAATGA	ATCGGCCAAC	8550
60	GCGCGGGGAG	AGGCGGTTTG	CGTATTGGGC	GCTCTTCCGC	TTCTCTGCTC	8600
	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA	8650



CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA 8700  
 AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC 8750  
 GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG AGCATCACAA 8800  
 5 AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT 8850  
 ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TCGCTCTCC GTTCCGACC 8900  
 CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC 8950  
 GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTT 9000  
 GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC 9050  
 10 GCCTTATCCG GTAACATATG TCTTGAGTCC AACCCGGTAA GACACGACTT 9100  
 ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG 9150  
 TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT 9200  
 AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG 9250  
 AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG 9300  
 GTGGTTTTTT TGTGTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT 9350  
 15 CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGAACGA 9400  
 AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA 9450  
 CCTAGATCCT TTTAAATTAA AAATGAAGTT TAAATCAAT CTAAAGTATA 9500  
 TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC 9550  
 TATCTCAGCG ATCTGTCTAT TTCGTTATC CATAGTTGCC TGAATCCCCG 9600  
 20 TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCAGTGCT 9650  
 GCAATGATAC CGCGAGACCC ACGCTCACC GCTCCAGATT TATCAGCAAT 9700  
 AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT 9750  
 CCGCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG AGTAAGTAGT 9800  
 TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT 9850  
 25 GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCTAAC 9900  
 GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAA AGCGGTTAGC 9950  
 TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTATC 10000  
 ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG 10050  
 TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA 10100  
 30 TAGTGTATGC GCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA 10150  
 TACCGCGCCA CATAGCAGAA CTTTAAAGT GCTCATCATT GGAAAACGTT 10200  
 CTTCCGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG 10250  
 ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC 10300  
 CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAATGCC GCAAAAAGG 10350  
 35 GAATAAGGGC GACACGAAA TGTGAATAC TCATACTCTT CCTTTTCAA 10400  
 TATTATTGAA GCATTATCA GGGTTATTGT CTCATGAGCG GATACATATT 10450  
 TGAATGTATT TAGAAAAATA AACAAATAGG GGTCCGCGC ACATTTCCCC 10500  
 GAAAAGTGCC AC 10512

40

SEQ ID NO:43 (pTnMOD (CMV-CHOVg-ent-ProInsulin-synPA))

1 ctgacgcgcc ctgtagcgcc gcattaagcg cggcggggtgt ggtgggttacg cgcagcgtga  
 61 ccgctacact tgccagegcc ctagegccc ctcctttcgc tttcttccct tcctttctcg  
 121 ccacgttcgc eggcatcaga ttggctattg gccattgcat acgttgatc catatcataa  
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
 241 tagttattaa tagtaataa ttacggggtc attagttcat agcccatata tggagttccg  
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt  
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
 421 atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc  
 481 aagtagccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
 541 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac  
 601 catggtgatg cggttttggc agtacatcaa tgggctgga tagcggtttg actcacgggg  
 661 atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg  
 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggag gtaggcgtgt  
 781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg  
 841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcgccg  
 901 ggaacgggag attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag  
 961 actctatagg cacacccctt tggtcttat gcattgctata ctgttttttg cttggggcct  
 1021 atacaccccc gcttcttat gctatagggtg atggatagc ttagcctata ggtgtgggtt  
 1081 attgaccatt attgaccact cccctatttg tgacgatact ttccattact aatccataac  
 1141 atggctcttt gccacaacta tctctatttg ctatatgcca atactctgtc cttcagagac  
 1201 tgacacggac tctgtatctt tacaggatgg ggtcccatct attatttaca aattcacata



5	1261	tacaacaacg	ccgtcccccg	tgcccgcagt	ttttattaaa	catagcgtgg	gatctccacg
	1321	cgaatctcgg	gtacgtgttc	cggacatggg	ctcttctccg	gtagcggcgg	agcttccaca
	1381	tccgagccct	gggtcccatgc	ctccagcggc	tcattggtcgc	tcggcagctc	cttgctccta
	1441	acagtggagg	ccagacttag	gcacagcaca	atgcccacca	ccaccagtgt	gccgcacaag
	1501	gccgtggcgg	taggggtatgt	gtctgaaaat	gagcgtggag	attgggctcg	cacggctgac
	1561	gcagatggaa	gacttaaggc	agcggcagaa	gaagatgcag	gcagctgagt	tgttgatttc
	1621	tgataagagt	cagaggtaac	tcccggttgcg	gtgctgttaa	cgggtggagg	cagtgtagtc
	1681	tgagcagtac	tcgttgctgc	cgcgcgcgcc	accagacata	atagctgaca	gactaacaga
10	1741	ctgttccttt	ccatgggtct	tttctgcagt	caccgtcgga	ccatgtgtga	acttgatatt
	1801	ttacatgatt	ctctttacca	attctgcccc	gaattacact	taaaacgact	caacagctta
	1861	acgttggtct	gccacgcatt	acttgactgt	aaaactctca	ctcttaccga	acttggccgt
	1921	aacctgccaa	ccaaagcgag	aacaaaacat	aacatcaaac	gaatcgaccg	attggttaggt
	1981	aatcgtcacc	tccacaaaga	gcgactcgct	gtataccggt	ggcatgctag	ctttatctgt
15	2041	tcgggcaata	cgatgcccac	tgtacttggt	gactggctctg	atattcgtga	gcaaaaacga
	2101	cttatgggat	tcgagcttc	agtcgcacta	cacggctcgt	ctgttactct	ttatgagaaa
	2161	gcgttcccg	tttcagagca	atgttcaaag	aaagctcatg	accaatttct	agccgacctt
	2221	gcgagcattc	taccgagtaa	caccacaccg	ctcattgtca	gtgatgctgg	ctttaaagtg
	2281	ccatgggata	aatccgttga	gaagctgggt	tggtactggg	taagtgcagt	aagaggaaaa
20	2341	gtacaatatg	cagacctagg	agcggaaaac	tggaaaaccta	tcagcaactt	acatgatatg
	2401	tcactagtc	actcaaagac	tttaggctat	aagaggctga	ctaaaagcaa	tccaatctca
	2461	tgccaaattc	tattgtataa	atctcgctct	aaaggccgaa	aaaatcagcg	ctcgacacgg
	2521	actcattgtc	accacccgtc	acctaaaatc	tactcagcgt	cggcaaaagg	gccatgggtt
	2581	ctagcaacta	acttacctgt	tgaaattcga	acacccaaac	aacttggtta	tatctattcg
25	2641	aagcgaatgc	agattgaaga	aaccttccga	gacttgaaaa	gtcctgccta	cggactaggg
	2701	ctacgccata	gccgaacgag	cagctcagag	cgttttgata	tcattgctgt	aatcgccctg
	2761	atgcttcaac	taacatgttg	gcttgccggc	gttcattgctc	agaaacaagg	ttgggacaag
	2821	cacttccagg	ctaacacagt	cagaaatcga	aacgtactct	caacagttcg	cttaggcattg
	2881	gaagttttgc	ggcattcttg	ctacacaata	acaagggaag	acttactcgt	ggctgcaacc
30	2941	ctactagctc	aaaatttatt	cacacatggg	tacgctttgg	ggaaattatg	ataatgatcc
	3001	agatcacttc	tggtctaata	aagatcagag	ctctagagat	ctgtgtgttg	gttttttgtg
	3061	gatctgctgt	gccttctagt	tgccagccat	ctgttggttg	ccccccccc	gtgccttctc
	3121	tgaccctgga	aggtgccact	cccactgtcc	tttccctaata	aaatgaggaa	attgcatcgc
	3181	attgtctgag	taggtgtcat	tctattctgg	gggggtgggg	ggggcagcac	agcaaggggg
35	3241	aggattggga	agacaatagc	aggcatgctg	gggatgcggt	gggtctctatg	ggtacctctc
	3301	tctctctctc	tctctctctc	tctctctctc	tctctcggtta	cctctctctc	tctctctctc
	3361	tctctctctc	tctctctctc	tccgtaccag	gtgctgaaga	attgaccggg	tgaccaaaagg
	3421	tgccctttat	catcacttta	aaaataaaaa	acaattactc	agtgcctgtt	ataagcagca
	3481	attaattatg	attgatgcct	acatcacaac	aaaaactgat	ttaacaaatg	gttggtctgc
40	3541	cttagaaaag	attattgaac	attatcttga	ttatattatt	gataataata	aaaaccttat
	3601	ccctatccaa	gaagtgtatg	ctatcattgg	ttggaatgaa	cttgaaaaaa	attagccttg
	3661	aatacattac	tggttaaggta	aacgccattg	tcagcaaaat	gatccaagag	aaccaactta
	3721	aagctttcct	gacggaatgt	taattctcgt	tgaccctgag	cactgatgaa	tcccctaattg
	3781	attttggtta	aaatcattaa	gttaagggtg	atacacatct	tgtcatatga	tcccggtaat
45	3841	gtgagtttag	tcactcatta	ggcaccaccg	gctttacact	ttatgcttcc	ggctcgtatg
	3901	ttgtgtggaa	ttgtgagcgg	ataacaattt	cacacaggaa	acagctatga	ccatgattac
	3961	gccaagcgcg	caattaaccc	tcactaaagg	gaacaaaagc	tgagctcca	ccgcggtggc
	4021	ggccgctcta	gaactagtgg	atcccccggg	catcagattg	gctattggcc	attgcatacg
	4081	ttgtatccat	atcataatat	gtacatttat	attggctcat	gtccaacatt	accgccatgt
50	4141	tgacattgat	tattgactag	ttattaatag	taatcaatta	cgggggtcatt	agttcatagc
	4201	ccatatatgg	agttcccgct	tacataactt	acggtaaatg	gcccgcctgg	ctgaccgccc
	4261	aacgaccccc	gccattgac	gtcaataatg	acgtatgttc	ccatagtaac	gccaataggg
	4321	actttccatt	gacgtcaatg	gggtggagtat	ttacggtaaa	ctgcccactt	ggcagtacat
	4381	caagtgtatc	atatgccaa	tacgccccct	attgacgtca	atgacggtaa	atggccccgc
55	4441	tggtattatg	cccagtacat	gaccttatgg	gactttccta	cttggcagta	catctacgta
	4501	ttagtcatcg	ctattaccat	gggtgatgctg	ttttggcagt	acatcaatgg	gcgtggatag
	4561	cggtttgact	cacggggatt	tccaagtctc	caccccatgg	acgtcaatgg	gagtttgttt
	4621	tggtacaaaa	atcaacggga	ctttccaaaa	tgctgtaaca	actccgcccc	attgacgcaa
	4681	atgggcggta	ggcgtgtacg	gtgggaggtc	tatataagca	gagctcgttt	agtgaaccgt
60	4741	cagatcgcc	ggagacgcca	tccacgctgt	tttgacctcc	atagaagaca	ccgggaccga
	4801	tccagcctcc	gcggacggga	acgggtgcat	ggaacgcgga	ttccccgtgc	caagagtgc
	4861	gtaagtaccg	cctatgact	ctataggcac	acccctttgg	ctcttatgca	tgctatactg
	4921	tttttggtct	ggggccctata	cacccccgct	tccttatgct	atagggtgatg	gtatagctta
	4981	gcctataggt	gtgggttatt	gaccattatt	gaccactccc	ctattgggtga	cgatacttcc
65	5041	cattactaat	ccataacatg	gctctttgcc	acaactatct	ctattggcta	tatgccaata
	5101	ctctgtcctt	cagagactga	cacggactct	gtatttttac	aggatggggg	cccatttatt
	5161	atttacaac	tcacatatac	aacaacgcgg	tcccccgctg	ccgcagtttt	tattaaacat
	5221	agcgtgggat	ctccacgcga	atctcgggta	cgtgttccgg	acatgggctc	ttctccggta
	5281	goggcggagc	ttccacatcc	gagccctggg	cccatgcctc	cagcggctca	tggtcgtcgc

5341 gcagctcctt gctcctaaca gtggaggcca gacttaggca cagcacaatg cccaccacca  
5401 ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgtc tgaaaatgag cgtggagatt  
5461 gggctcgcac ggctgacgca gatggaagac ttaaggcagc ggcagaagaa gatgcaggca  
5521 gctgagttgt tgtattctga taagagtcag aggtaactcc cgttgcggtg ctgttaacgg  
5581 tggagggcag tgtagtctga gcagtactcg ttgctgccgc gcgcgccacc agacataata  
5641 gctgacagac taacagactg ttcccttcca tgggtctttt ctgcagtcac cgtcggatca  
5701 atgggctcca tcggtgcagc aagcatggaa ttttgttttg atgtattcaa ggagctcaaa  
5761 gtccaccatg ccaatgagaa catcttctac tgccccattg ccatcatgtc agctctagcc  
5821 atgggtatacc tgggtgcaaa agacagcacc aggcacacaaa taaataagggt tgttcgcttt  
5881 gataaacttc caggattcgg agacagtatt gaagctcagt gtggcacatc tgtaaacggt  
5941 cactcttcac ttagagacat cctcaaccaa atcaccaaac caaatgatgt ttattcgttc  
6001 agccttgcca gtagacttta tggctgaagag agatacccaa tccctgccaga atacttgtag  
6061 tgtgtgaagg aactgtatag aggaggcttg gaacctatca actttcaaac agctgcagat  
6121 caagccagag agctcatcaa ttccctgggtg gaaagtcaga caaatggaat tatcagaaat  
6181 gtccttcagc caagctccgt ggattctcaa actgcaatgg ttctgggtta tgccattgtc  
6241 ttcaaaggac tgtgggagaa agcatttaag gatgaagaca cacaagcaat gcctttcaga  
6301 gtgactgagc aagaaaagcaa acctgtgcag atgatgtacc agattgggtt atttagagtg  
6361 gcatcaatgg cttctgagaa aatgaagatc ctggagcttc catttgccag tgggacaatg  
6421 agcatgttgg tgctgttgcc tgatgaagtc tcaggccttg agcagcttga gagtataatc  
6481 aactttgaaa aactgactga atggaccagt tctaattgta tgggaagagag aagatcaaa  
6541 tgtacttacc tcgcatgaag atggaggaaa aatacaacct cacatctgtc ttaatggcta  
6601 tgggcattac tgacgtgttt agctcttcag ccaatctgtc tggcatctcc tcagcagaga  
6661 gcctgaagat atctcaagct gtccatgcag cacatgcaga aatcaatgaa gcaggcagag  
6721 aggtggtagg gtcagcagag gctggagtgg atgctgcaag cgtctctgaa gaatttaggg  
6781 ctgaccatcc attcctcttc tgtatcaagc acatcgcaac caacgccgtt ctctctcttt  
6841 ggcagatgtg ttcccgcg gcagcagatg acgcaccagc agatgacgca ccagcagatg  
6901 acgcaccagc agatgacgca ccagcagatg acgcaacaac atgtatcctg aaaggctctt  
6961 gtggttggtg cggcctgctg gatgacgatg acaaatgtgt gaaccaacac ctgtgcggct  
7021 cacacctggg ggaagctctc tactagtgt gcggggaacg aggcctcttc tacacacca  
7081 agaccgccc ggaggcagag gacctgcagg tggggcagggt ggagctgggc gggggccctg  
7141 gtgcaggcag cctgcagccc ttggcccttg aggggtccct gcagaagcgt ggcattgtgg  
7201 aacaatgctg taccagcatc tgctccctct accagctgga gaactactgc aactaggggc  
7261 cctaaagggc gaattatcgc ggccgctcta gaccaggcgc ctggatccag atcactctg  
7321 gctaataaaa gatcagagct ctgagatct gtgtgttggg ttttgttggg tctgctgtgc  
7381 cttctagttg ccagccatct gttgtttgcc cctccccctg gccttccttg accctggaag  
7441 gtgccactcc cactgtcctt tcctaataaa atgaggaaat tgcacgcat tgtctgagta  
7501 ggtgtcattc tattctgggg ggtgggggtg ggcagcacag caagggggag gattgggaag  
7561 acaatagcag gcatgctggg gatgcgggtg gctctatggg tactctctc tctctctc  
7621 tctctctcac tctctctc tctcgttacc tctcctcgag ggggggcccgt gtaccaatt  
7681 cgccctatag tgagtcgtat tacgcgcgct cactggcgt cgttttaca cgtcgtgact  
7741 gggaaaaccc tggcgttacc caatttaac gccttgacgc acatccccct ttgcgcagct  
7801 ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg  
7861 gcgaatggaa attgtaagcg ttaatatatt gttaaaatc gcgttaaatt tttgttaa  
7921 cagctcattt tttaaccaat aggcggaat cggcaaaatc ccttataaat caaagaata  
7981 gaccgagata ggggtgagtg ttgttccagt ttggaacaag agtccactat taaagaacgt  
8041 ggactccaac gtcaaaggcc gaaaaaccgt ctatcagggc gatggcccac tactccggga  
8101 tcatatgaca agatgtgtat ccaccttaac ttaatgattt ttaccaaata cattagggga  
8161 ttcacagtg ctcagggtca acgagaatta acattccgtc aggaagcctt atgatgatga  
8221 tgtgcttaaa aacttactca atggctgggt atgcatatcg caatacatgc gaaaaacct  
8281 aaagagcttg ccgataaaaa aggcgaattt attgctattt accgcggctt ttatttgagc  
8341 ttgaaagata aataaaatag atagggttta tttgaagcta aatcttcttt atcgtaaaaa  
8401 atgcccctct ggggttatcaa gagggtcatt atatttcgcg gaataacatc atttgggtgac  
8461 gaaataacta agcacttgct tcctgtttac tccccgagc ttgaggggtt aacatgaagg  
8521 tcatcgatag caggataata atacagtaaa acgctaaacc aataatccaa atocagccat  
8581 cccaaattgg tagtgaatga ttataaataa cagcaaacag taatgggcca ataacaccgg  
8641 ttgcattggg aaggetcacc aataatccct gtaaagcacc ttgctgatga ctctttgttt  
8701 ggatagacat cactccctgt aatgcaggta aagcgatccc accaccagcc aataaaatta  
8761 aaacagggaa aactaaccaa ccttcagata taaacgctaa aaaggcaaat gcactactat  
8821 ctgcaataaa tccgagcagt actgcccgtt ttogcccat ttagtggcta ttcttctgc  
8881 cacaaggct tgggaatactg agtgtaaaag accaagacc gtaatgaaaa gccaaaccatc  
8941 atgctattca tcatcacgat ttctgttaata gcaccacacc gtgctggatt ggctatcaat  
9001 gcgctgaaat aataatcaac aaatggcatc gttaaataag tgatgtatac cgatcagctt  
9061 ttgttccctt tagtgagggt taattgcgcg cttggcgtaa tcatgggtcat agctgtttcc  
9121 tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagt  
9181 taaagcctgg ggtgcctaata gagtgaagta actcacatta attgcgttgc gctcactgcc  
9241 cgctttccag tccggaaacc tgcgtgcca gctgcattaa tgaatcgccc aacgcgcggg  
9301 gagaggcggg ttgcgtattg ggcgctcttc cgttccctcg ctcactgact cgtgcgctc  
9361 ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag gcggtaatac ggttatccac

5 9421 agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa  
 9481 ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cggccccctg acgagcatca  
 9541 caaaaatcga cgctcaagtc agaggtggcg aaaccgcaca ggactataaa gataccaggc  
 9601 gttttccccc ggaagctccc tcgtgcgctc tcctgttccg accctgcgcg ttaccggata  
 9661 cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta  
 9721 tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac ccccggttca  
 9781 gcccagaccgc tgcgccttat ccggttaacta tcgtcttgag tccaaccggy taagacacga  
 9841 cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg  
 10 9901 tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg  
 9961 tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg  
 10021 caaacaacc accgctggta ggcgtgggtt ttttgttgc aagcagcaga ttaccgcgag  
 10081 aaaaaaagga tctcaagaag atcctttgat cttttctacg ggggtctgacg ctcaagtggaa  
 10141 cgaaaactca cgtaagggga ttttgggtcat gagattatca aaaaggatct tcacctagat  
 10201 ctttttaaat taaaaatgaa gtttttaate aatctaaagt atatatgagt aaacttggte  
 15 10261 tgacagttac caatgcttaa tcagtggagc acctatctca gcgatctgtc tatttcgttc  
 10321 atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc  
 10381 tggccccagt gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcagc  
 10441 aataaaccag ccagccggaa gggccgagcg cagaagtggc cctgcaactt tatccgcctc  
 10501 catccagtct attaatgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt  
 20 10561 gcgcaacgtt gttgccattg ctacaggcat cgtgggtgta cgtcgtcgtc ttggtatggc  
 10621 ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatcccca tgttgtgcaa  
 10681 aaaagcgggt agtccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt  
 10741 atcactcatg gttatggcag cactgcataa ttctcttact gtcagccat ccgtaagatg  
 10801 cttttctgtg actggtgagt actcaacca gtcattctga gaatagtga tgcggcgacc  
 25 10861 gagttgctct tgcggcggt caatacggga taataccgcy ccacatagca gaactttaaa  
 10921 agtgctcatc attggaatac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt  
 10981 gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt  
 11041 caccagcgtt tctgggtgag caaaaacagg aaggcaaat gccgcaaaaa agggataaag  
 30 11101 ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta  
 11161 tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaat  
 11221 aggggttccg cgcacatttc cccgaaaagt gccac

35 SEQ ID NO:44 (pTnMod(Oval/ENT tag/Proins/PA) - QUAIL)  
 CTGACGCGCC CTGTAGCGGC GCATTAAAGCG CGGCGGGTGT GGTGGTTACG 50  
 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC 100  
 TTTCTTCCCT TCCTTTCTCG CCACGTTTCG CGGCATCAGA TTGGCTATTG 150  
 GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200  
 CATGTCCAAC ATTACCGCCA TGTTGACATT GATTATTGAC TAGTTATTAA 250  
 40 TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300  
 CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCAACGACC 350  
 CCCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400  
 GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAAGTGCCCA 450  
 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC CCTATTGACG 500  
 45 TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA 550  
 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600  
 CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG 650  
 ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700  
 TTTTGGCACC AAAATCAACG GGAATTTCCA AAATGTCGTA ACAACTCCGC 750  
 50 CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 800  
 GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850  
 TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900  
 GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA 950  
 CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000  
 55 CTGTTTTTGG CTTGGGGCCT ATACACCCCG GCTTCCTTAT GCTATAGGTG 1050  
 ATGGTATAGC TTAGCCTATA GGTGTGGGT ATTGACCATT ATTGACCACT 1100  
 CCCCTATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150  
 GCCACAATA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTGAGAGAC 1200  
 TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATTATTTACA 1250  
 60 AATTCACATA TACAACAACG CCGTCCCCCG TGCCCGCAGT TTTTATTAAA 1300  
 CATAGCGTGG GATCTCCACG CGAATCTCGG GTACGTGTTT CCGACATGGG 1350  
 CTCTTCTCCG GTAGCGGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400  
 CTCCAGCGGC TCATGGTTCG TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450  
 CCAGACTTAG GCACAGCACA ATGCCACCA CCACAGTGT GCCGCACAAG 1500



GCCGTGGCGG TAGGGTATGT GTCTGAAAAT GAGCGTGGAG ATTGGGCTCG 1550  
CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600  
GCAGCTGAGT TGTGTATTTC TGATAAGAGT CAGAGGTAAC TCCCCTGCG 1650  
GTGCTGTTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700  
5 CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCCCTT 1750  
CCATGGGTCT TTTCTGCAGT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800  
TTACATGATT CTCTTTACCA ATTCTGCCCC GAATTACACT TAAAACGACT 1850  
CAACAGCTTA ACGTTGGCTT GCCACGCATT ACTTGACTGT AAAACTCTCA 1900  
CTCTTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCGAG AACAAAACAT 1950  
10 AACATCAAAC GAATCGACCG ATTGTTAGGT AATCGTCACC TCCACAAAGA 2000  
GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050  
GATGCCCATT GTACTTGTG ACTGGTCTGA TATTCGTGAG CAAAAACGAC 2100  
TTATGGTATT GCGAGCTTCA GTCGCACCTAC ACGGTCGTTT TGTACTCTT 2150  
TATGAGAAAG CGTTCGCGCT TTCAGAGCAA TGTTCAAAGA AAGCTCATGA 2200  
15 CCAATTTCTA GCCGACCTTG CGAGCATTCT ACCGAGTAAC ACCACACCGC 2250  
TCATTGTCAG TGATGCTGGC TTTAAAGTGC CATGGTATAA ATCCGTTGAG 2300  
AAGCTGGGTT GGTACTGGTT AAGTCGAGTA AGAGGAAAAG TACAATATGC 2350  
AGACCTAGGA GCGGAAAAC GGAACCTAT CAGCAACTTA CATGATATGT 2400  
CATCTAGTCA CTCAAAGACT TTAGGCTATA AGAGGCTGAC TAAAAGCAAT 2450  
20 CCAATCTCAT GCCAAATCT ATTGTATAAA TCTCGCTCTA AAGGCCGAAA 2500  
AAATCAGCGC TCGACACGGA CTCATTGTCA CCACCCGTCA CCTAAAATCT 2550  
ACTCAGCGTC GGCAAAGGAG CCATGGGTTT TAGCAACTAA CTTACCTGTT 2600  
GAAATTCGAA CACCCAAACA ACTTGTTAAT ATCTATTCTGA AGCGAATGCA 2650  
GATTGAAGAA ACCTTCCGAG ACTTGAAAAG TCCTGCCTAC GGAAGTAGGC 2700  
25 TACGCCATAG CCGAACGAGC AGCTCAGAGC GTTTTGATAT CATGCTGCTA 2750  
ATCGCCCTGA TGCTTCAACT AACATGTTGG CTTGCGGGCG TTCATGCTCA 2800  
GAAACAAGGT TGGGACAAGC ACTTCCAGGC TAACACAGTC AGAAATCGAA 2850  
ACGTACTCTC AACAGTTCGC TTAGGCATGG AAGTTTTCG GCATTCTGGC 2900  
TACACAATAA CAAGGGAAGA CTTACTCGTG GCTGCAACCC TACTAGCTCA 2950  
30 AAATTTATTC ACACATGGTT ACGCTTGGG GAAATTATGA TAATGATCCA 3000  
GATCACTTCT GGCTAATAAA AGATCAGAGC TCTAGAGATC TGTGTGTTGG 3050  
TTTTTTGTGG ATCTGCTGTG CTTCTAGTT GCCAGCCATC TGTGTTTGC 3100  
CCCTCCCCCG TGCCTTCCTT GACCTGGAA GGTGCCACTC CCCTGTCCT 3150  
TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT 3200  
35 CTATTCTGGG GGGTGGGGTG GGCAGCACA GCAAGGGGGA GGATTGGGAA 3250  
GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCTCTCT 3300  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCGGTAC CTCTCTCTCT 3350  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CGGTACCAGG TGCTGAAGAA 3400  
TTGACCCGGT GACCAAAGGT GCCTTTTATC ATCACTTTAA AAATAAAAAA 3450  
40 CAATTACTCA GTGCCTGTTA TAAGCAGCAA TTAATTATGA TTGATGCCTA 3500  
CATCACAACA AAAACTGATT TAACAAATGG TTGGTCTGCC TTAGAAAGTA 3550  
TATTTGAACA TTATCTTGAT TATATTATG ATAATAATAA AAACCTTATC 3600  
CCTATCCAAG AAGTGATGCC TATCATTGGT TGGAATGAAC TTGAAAAAAA 3650  
TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATTG 3700  
45 ATCCAAGAGA ACCAACTTAA AGCTTTCCTG ACGGAATGTT AATTCTCGTT 3750  
GACCCTGAGC ACTGATGAAT CCCCTAATGA TTTTGGTAAA AATCATTAA 3800  
TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850  
CACTCATTAG GCACCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT 3900  
TGTGTGGAAT TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC 3950  
50 CATGATTACG CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT 4000  
GGAGCTCCAC CGCGGTGGCG GCCGCTCTAG AACTAGTGGA TCCCCGGGG 4050  
AGGTCAGAAT GGTTCCTTTA CTGTTTGTCA ATTCTATTAT TTCAATACAG 4100  
AACAAAAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG 4150  
TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTACAA TTCTCTGTC 4200  
55 ATCTGCCAGG CTGGAAGATC ATGGAAGATC TCTGAGGAAC ATTGCAAGTT 4250  
CATACCATAA ACTCATTTGG AATTGAGTAT TATTTTGCTT TGAATGGAGC 4300  
TATGTTTTGC AGTCCCTCA GAAGAAAAGC TTGTTATAAA GCGTCTACAC 4350  
CCATCAAAG ATATATTTAA ATATTCCAAC TACAGAAAGA TTTTGTCTGC 4400  
TCTTCACTCT GATCTCAGTT GGTTTCTTCA CGTACATGCT TCTTTATTTG 4450  
60 CCTATTTTGT CAAGAAAATA ATAGGTCAAG TCCTGTTCTC ACTTATCTCC 4500  
TGCCTAGCAT GGCTTAGATG CACGTTGTAC ATCAAGAAG GATCAAATGA 4550

	AACAGACTTC	TGGTCTGTTA	CAACAACCAT	AGTAATAAAC	AGACTAACTA	4600
	ATAATTGCTA	ATTATGTTTT	CCATCTCTAA	GGTTCCCACA	TTTTTCTGTT	4650
	TTAAGATCCC	ATTATCTGGT	TGTAAGTGAA	GCTCAATGGA	ACATGAACAG	4700
	TATTTCTCAG	TCTTTTCTCC	AGCAATCCTG	ACGGATTAGA	AGAACTGGCA	4750
5	GAAAACACTT	TGTTACCCAG	AATTAAAAAC	TAATATTTGC	TCTCCCTTCA	4800
	ATCCAAAATG	GACCTATTGA	AACTAAAATC	TGACCCAATC	CCATTAAATT	4850
	ATTTCTATGG	CGTCAAAGGT	CAAACTTTGT	AAGGGAACCT	GTGGGTGGGT	4900
	CCCAATTCAG	GCTATATATT	CCCCAGGGCT	CAGCCAGTGG	ATCCATGGGC	4950
	TCCATCGGTG	CAGCAAGCAT	GGAATTTTGT	TTTGATGTAT	TCAAGGAGCT	5000
10	CAAAGTCCAC	CATGCCAATG	ACAACATGCT	CTACTCCCCC	TTTGCCATCT	5050
	TGTCAACTCT	GGCCATGGTC	TTCTAGGTG	CAAAAGACAG	CACCAGGACC	5100
	CAGATAAATA	AGGTGTGTTCA	CTTTGATAAA	CTTCCAGGAT	TCGGAGACAG	5150
	TATTGAAGCT	CAGTGTGGCA	CATCTGTAAA	TGTTCACTCT	TCACTTAGAG	5200
	ACATACTCAA	CCAAATCACC	AAACAAAATG	ATGCTTATTC	GTTCAGCCTT	5250
15	GCCAGTAGAC	TTTATGCTCA	AGAGACATAC	ACAGTCGTGC	CGGAATACTT	5300
	GCAATGTGTG	AAGGAACTGT	ATAGAGGAGG	CTTAGAATCC	GTCAACTTTC	5350
	AAACAGCTGC	AGATCAAGCC	AGAGGCCTCA	TCAATGCCTG	GGTAGAAAGT	5400
	CAGACAAACG	GAATTATCAG	AAACATCCTT	CAGCCAAGCT	CCGTGGATTG	5450
	TCAAACCTGCA	ATGGTCCTGG	TTAATGCCAT	TGCCTTCAAG	GGACTGTGGG	5500
20	AGAAAGCATT	TAAGGCTGAA	GACACGCAAA	CAATACCTTT	CAGAGTGACT	5550
	GAGCAAGAAA	GCAAACCTGT	GCAGATGATG	TACCAGATTG	GTTCATTTTAA	5600
	AGTGGCATCA	ATGGCTTCTG	AGAAAATGAA	GATCCTGGAG	CTTCCATTTG	5650
	CCAGTGGAAC	AATGAGCATG	TTGGTGCTGT	TGCCTGATGA	TGTCTCAGGC	5700
	CTTGAGCAGC	TTGAGAGTAT	AATCAGCTTT	GAAAACTGA	CTGAATGGAC	5750
25	CAGTTCTAGT	ATTATGGAAG	AGAGGAAGGT	CAAAGTGTAC	TTACCTCGCA	5800
	TGAAGATGGA	GGAGAAATAC	AACCTCACAT	CTCTCTTAAT	GGCTATGGGA	5850
	ATTACTGACC	TGTTCAAGCT	TTCAAGCAAT	CTGTCTGGCA	TCTCCTCAGT	5900
	AGGGAGCCTG	AAGATATCTC	AAGCTGTCCA	TGCAGCACAT	GCAGAAATCA	5950
	ATGAAGCGGG	CAGAGATGTG	GTAGGCTCAG	CAGAGGCTGG	AGTGGATGCT	6000
30	ACTGAAGAAT	TTAGGGCTGA	CCATCCATTG	CTCTTCTGTG	TCAAGCACAT	6050
	CGAAACCAAC	GCCATTCTCC	TCTTTGGCAG	ATGTGTTTCT	CCGCGGCCAG	6100
	CAGATGACGC	ACCAGCAGAT	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	6150
	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	GACGCAACAA	CATGTATCCT	6200
	GAAAGGCTCT	TGTGGCTGGA	TCGGCCTGCT	GGATGACGAT	GACAAATTTG	6250
35	TGAACCAACA	CCTGTGCGGC	TCACACCTGG	TGGAAGCTCT	CTACCTAGTG	6300
	TGCGGGGAAC	GAGGCTTCTT	CTACACACCC	AAGACCCGCC	GGGAGGCAGA	6350
	GGACCTGCAG	GTGGGGCAGG	TGGAGCTGGG	CGGGGGCCCT	GGTGCAGGCA	6400
	GCCTGCAGCC	CTTGGCCCTG	GAGGGGTCCC	TGCAGAAGCG	TGGCATTGTG	6450
	GAACAATGCT	GTACCAGCAT	CTGCTCCCTC	TACCAGCTGG	AGAACTACTG	6500
40	CAACTAGGGC	GCCTGGATCC	AGATCACTTC	TGGCTAATAA	AAGATCAGAG	6550
	CTCTAGAGAT	CTGTGTGTTG	GTTTTTTGTG	GATCTGCTGT	GCCTTCTAGT	6600
	TGCCAGCCAT	CTGTTGTTTG	CCCCTCCCCC	GTGCCTTCCT	TGACCCTGGA	6650
	AGGTGCCACT	CCCCTGTCC	TTTCCTAATA	AAATGAGGAA	ATTGCATCGC	6700
	ATTGTCTGAG	TAGGTGTCAT	TCTATTCTGG	GGGGTGGGGT	GGGCAGCAC	6750
45	AGCAAGGGGG	AGGATTGGGA	AGACAATAGC	AGGCATGCTG	GGGATGCGGT	6800
	GGGCTCTATG	GGTACCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	6850
	TCTCTCGGTA	CCTCTCTCGA	GGGGGGGCCC	GGTACCCAAT	TCGCCCTATA	6900
	GTGAGTCGTA	TTACGCGCGC	TCAGTGGCCG	TCGTTTTCAC	ACGTCGTGAC	6950
	TGGGAAAACC	CTGGCGTTAC	CCAACCTAAT	CGCCTTGCAG	CACATCCCCC	7000
50	TTTCGCCAGC	TGGCGTAATA	GCGAAGAGGC	CCGCACCGAT	CGCCCTTCCC	7050
	AACAGTTGCG	CAGCCTGAAT	GGCGAATGGA	AATTGTAAGC	GTTAATATTT	7100
	TGTTAAAATT	CGCGTTAAAT	TTTTGTAAAT	TCAGCTCATT	TTTTAACCAA	7150
	TAGGCCGAAA	TCGGCAAAAT	CCCTTATAAA	TCAAAAGAAT	AGACCGAGAT	7200
	AGGGTTGAGT	GTTGTTCCAG	TTTGGAAACAA	GAGTCCACTA	TTAAAGAACG	7250
55	TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCTATCAGGG	CGATGGCCCA	7300
	CTACTCCGGG	ATCATATGAC	AAGATGTGTA	TCCACCTTAA	CTTAATGATT	7350
	TTTACCAAAA	TCATTAGGGG	ATTCATCAGT	GCTCAGGGTC	AACGAGAATT	7400
	AACATTCCGT	CAGGAAAGCT	TATGATGATG	ATGTGCTTAA	AAACTTACTC	7450
	AATGGCTGGT	TATGCATATC	GCAATACATG	CGAAAAACCT	AAAAGAGCTT	7500
60	GCCGATAAAA	AAGGCCAATT	TATTGCTATT	TACCGCGGCT	TTTTATTGAG	7550
	CTTGAAAGAT	AAATAAAATA	GATAGGTTTT	ATTTGAAGCT	AAATCTTCTT	7600

	TATCGTAAAA	AATGCCCTCT	TGGGTATCA	AGAGGGTCAT	TATATTTTCGC	7650
	GGAATAACAT	CATTGTTGA	CGAAATAACT	AAGCACTTGT	CTCCTGTTTA	7700
	CTCCCCTGAG	CTTGAGGGGT	TAACATGAAG	GTCATCGATA	GCAGGATAAT	7750
5	AATACAGTAA	AACGCTAAAC	CAATAATCCA	AATCCAGCCA	TCCCAAATTG	7800
	GTAGTGAATG	ATTATAAATA	ACAGCAAACA	GTAATGGGCC	AATAACACCG	7850
	GTTGCATTGG	TAAGGCTCAC	CAATAATCCC	TGTAAAGCAC	CTTGCTGATG	7900
	ACTCTTTGTT	TGGATAGACA	TCACTCCCTG	TAATGCAGGT	AAAGCGATCC	7950
	CACCACCAGC	CAATAAAATT	AAAACAGGGA	AAACTAACCA	ACCTTCAGAT	8000
	ATAAACGCTA	AAAAGGCAAA	TGCACTACTA	TCTGCAATAA	ATCCGAGCAG	8050
10	TACTGCCGTT	TTTTCGCCCC	ATTTAGTGGC	TATTCCTCCT	GCCACAAAGG	8100
	CTTGAATAC	TGAGTGTAAG	AGACCAAGAC	CCGCTAATGA	AAAGCCAACC	8150
	ATCATGCTAT	TCCATCCAAA	ACGATTTTCG	GTAAATAGCA	CCCACACCGT	8200
	TGCGGGAATT	TGGCCTATCA	ATTGCGCTGA	AAAATAAATA	ATCAACAAAA	8250
	TGGCATCGTT	TTAAATAAAG	TGATGTATAC	CGAATTCAGC	TTTTGTTCCC	8300
15	TTTAGTGAGG	GTTAATTGCG	CGCTTGGCGT	AATCATGGTC	ATAGCTGTTT	8350
	CCTGTGTGAA	ATTGTTATCC	GCTCACAATT	CCACACAACA	TACGAGCCGG	8400
	AAGCATAAAG	TGTAAAGCCT	GGGGTGCCTA	ATGAGTGAGC	TAATCACAT	8450
	TAATTGCGTT	GCGCTCACTG	CCCGCTTTCC	AGTCGGGAAA	CCTGTGCTGC	8500
	CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCGTAT	8550
20	TGGGCGCTCT	TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	TCGGTCTGTC	8600
	GGCTGCGGCG	AGCGGTATCA	GCTCACTCAA	AGGCGGTAAT	ACGGTTATCC	8650
	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	ATGTGAGCAA	AAGGCCAGCA	8700
	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	TTCCATAGGC	8750
	TCCGCCCCCC	TGACGAGCAT	CACAAAAATC	GACGCTCAAG	TCAGAGGTGG	8800
25	CGAAACCCGA	CAGGACTATA	AAGATAACCAG	GCGTTTCCCC	CTGGAAGCTC	8850
	CCTCGTGCGC	TCTCCTGTTT	CGACCCTGCC	GCTTACCGGA	TACCTGTCCG	8900
	CCTTTCTCCC	TTGCGGAAGC	GTGGCGCTTT	CTCATAGCTC	ACGCTGTAGG	8950
	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	AAGCTGGGCT	GTGTGCACGA	9000
	ACCCCCCGTT	CAGCCCGACC	GCTGCGCCTT	ATCCGGTAAC	TATCGTCTTG	9050
30	AGTCCAACCC	GGTAAGACAC	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	9100
	AACAGGATTA	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	9150
	GTGGTGGCCT	AACTACGGCT	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	9200
	CTCTGCTGAA	GCCAGTTACC	TTGCGAAAAA	GAGTTGGTAG	CTCTTGATCC	9250
	GGCAAACAAA	CCACCGCTGG	TAGCGGTGGT	TTTTTTGTTT	GCAAGCAGCA	9300
35	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	ATCTTTTCTA	9350
	CGGGGTCTGA	CGCTCAGTGG	AACGAAAAT	CACGTTAAGG	GATTTTGCTC	9400
	ATGAGATTAT	CAAAAAGGAT	CTTCACCTAG	ATCCTTTTAA	ATTAAAAATG	9450
	AAGTTTAA	TCAATCTAAA	GTATATATGA	GTAACTTGG	TCTGACAGTT	9500
	ACCAATGCTT	AATCAGTGAG	GCACCTATCT	CAGCGATCTG	TCTATTTCTG	9550
40	TCATCCATAG	TTGCCTGACT	CCCCGTGCTG	TAGATAACTA	CGATACGGGA	9600
	GGGCTTACCA	TCTGGCCCCA	GTGCTGCAAT	GATACCGCGA	GACCCACGCT	9650
	CACCGGCTCC	AGATTTATCA	GCAATAAAC	AGCCAGCCGG	AAGGGCCGAG	9700
	CGCAGAAGTG	GTCCTGCAAC	TTTATCCGCC	TCCATCCAGT	CTATTAATTG	9750
	TTGCCGGGAA	GCTAGAGTAA	GTAAGTCCGC	AGTTAATAGT	TTGCGCAACG	9800
45	TTGTTGCCAT	TGCTACAGGC	ATCGTGGTGT	CACGCTCGTC	GTTTGGTATG	9850
	GCTTCATTCA	GCTCCGGTTC	CCAACGATCA	AGGCGAGTTA	CATGATCCCC	9900
	CATGTTGTGC	AAAAAAGCGG	TTAGCTCCTT	CGGTCCTCCG	ATCGTTGTCA	9950
	GAAGTAAGTT	GGCCGCAGTG	TTATCACTCA	TGGTTATGGC	AGCACTGCAT	10000
	AATTCTCTTA	CTGTCATGCC	ATCCGTAAGA	TGCTTTTCTG	TGACTGGTGA	10050
50	GTACTCAACC	AAGTCATTCT	GAGAATAGTG	TATGCGGCGA	CCGAGTTGCT	10100
	CTTGCCCCGC	GTCAATACGG	GATAATACCG	CGCCACATAG	CAGAACTTTA	10150
	AAAGTGCTCA	TCATTGGAAA	ACGTTCTTCG	GGGCGAAAAC	TCTCAAGGAT	10200
	CTTACCGCTG	TTGAGATCCA	GTTGATGTA	ACCCACTCGT	GCACCCAACT	10250
	GATCTTCAGC	ATCTTTTACT	TTACCAGCG	TTCTGGGTG	AGCAAAAACA	10300
55	GGAAGGCAAA	ATGCCGCAAA	AAAGGGAATA	AGGGCGACAC	GGAAATGTTG	10350
	AATACTCATA	CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	10400
	ATTGTCTCAT	GAGCGGATAC	ATATTTGAAT	GTATTTAGAA	AAATAAACAA	10450
	ATAGGGGTTC	CGCGCACATT	TCCCCGAAA	GTGCCAC		10487
60	SEQ ID NO:45 (pTnMod(Oval/ENT tag/P146/PA) - Chicken)					



	CTGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	50
	CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCTTTCGC	100
	TTTCTTCCCT	TCCTTTCTCG	CCACGTTCGC	CGGCATCAGA	TTGGCTATTG	150
5	GCCATTGCAT	ACGTTGTATC	CATATCATAA	TATGTACATT	TATATTGGCT	200
	CATGTCCAAC	ATTACCGCCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	250
	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT	AGCCCATATA	TGGAGTTCCG	300
	CGTTACATAA	CTTACGGTAA	ATGGCCCCGC	TGGCTGACCG	CCCAACGACC	350
	CCCGCCCAT	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	400
10	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT	AAACTGCCCA	450
	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTACGCCC	CCTATTGACG	500
	TCAATGACGG	TAAATGGCCC	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA	550
	TGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC	600
	CATGGTGATG	CGGTTTGGC	AGTACATCAA	TGGGCGTGGA	TAGCGGTTTG	650
15	ACTCACGGGG	ATTTCCAAGT	CTCCACCCCA	TTGACGTCAA	TGGGAGTTTG	700
	TTTTGGCACC	AAAATCAACG	GGACTTTCCA	AAATGTCGTA	ACAACCTCCG	750
	CCCATTGACG	CAAATGGGCG	GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	800
	GCAGAGCTCG	TTTAGTGAAC	CGTCAGATCG	CCTGGAGACG	CCATCCACGC	850
	TGTTTTGACC	TCCATAGAAG	ACACCGGGAC	CGATCCAGCC	TCCGCGGCCG	900
20	GGAACGGTGC	ATTGGAACGC	GGATTCCCCG	TGCCAAGAGT	GACGTAAGTA	950
	CCGCCTATAG	ACTCTATAGG	CACACCCCTT	TGGCTCTTAT	GCATGCTATA	1000
	CTGTTTTTGG	CTTGGGGCCT	ATACACCCCC	GCTTCCTTAT	GCTATAGGTG	1050
	ATGGTATAGC	TTAGCCTATA	GGTGTGGGTT	ATTGACCATT	ATTGACCACT	1100
	CCCCTATTGG	TGACGATACT	TTCCATTACT	AATCCATAAC	ATGGCTCTTT	1150
25	GCCACAATA	TCTCTATTGG	CTATATGCCA	ATACTCTGTC	CTTCAGAGAC	1200
	TGACACGGAC	TCTGTATTTT	TACAGGATGG	GGTCCCATT	ATTATTTACA	1250
	AATTCACATA	TACAACAACG	CCGTCCCCCG	TGCCCCGAGT	TTTATTAAA	1300
	CATAGCGTGG	GATCTCCACG	CGAATCTCGG	GTACGTGTTT	CGGACATGGG	1350
	CTCTTCTCCG	GTAGCGGCGG	AGCTTCCACA	TCCGAGCCCT	GGTCCCATGC	1400
30	CTCCAGCGGC	TCATGGTCGC	TCGGCAGCTC	CTTGCTCCTA	ACAGTGGAGG	1450
	CCAGACTTAG	GCACAGCACA	ATGCCCACCA	CCACCAGTGT	GCCGCACAAG	1500
	GCCGTGGCGG	TAGGGTATGT	GTCTGAAAAT	GAGCGTGGAG	ATTGGGCTCG	1550
	CACGGCTGAC	GCAGATGGAA	GACTTAAGGC	AGCGGCAGAA	GAAGATGCAG	1600
	GCAGCTGAGT	TGTTGTATTC	TGATAAGAGT	CAGAGGTAAC	TCCCGTTGCG	1650
35	GTGCTGTTAA	CGGTGGAGGG	CAGTGTAGTC	TGAGCAGTAC	TCGTTGCTGC	1700
	CGCGCGCGCC	ACCAGACATA	ATAGCTGACA	GACTAACAGA	CTGTTCTTTT	1750
	CCATGGGTCT	TTTCTGCAGT	CACCGTCGGA	CCATGTGTGA	ACTTGATATT	1800
	TTACATGATT	CTCTTTACCA	ATTCTGCCCC	GAATTACACT	TAAAACGACT	1850
	CAACAGCTTA	ACGTTGGCTT	GCCACGCATT	ACTTGACTGT	AAAACCTCTCA	1900
40	CTCTTACCGA	ACTTGCCCGT	AACCTGCCAA	CAAAGCGAG	AACAAAACAT	1950
	AACATCAAAC	GAATCGACCG	ATTGTTAGGT	AATCGTCACC	TCCACAAAGA	2000
	GCGACTCGCT	GTATACCGTT	GGCATGCTAG	CTTTATCTGT	TCGGGAATAC	2050
	GATGCCCAT	GTACTTGTTG	ACTGGTCTGA	TATTCGTGAG	CAAAAACGAC	2100
	TTATGGTATT	GCGAGCTTCA	GTCGCACTAC	ACGGTCGTTT	TGTTACTCTT	2150
45	TATGAGAAAG	CGTTCGCGCT	TTGAGAGCAA	TGTTCAAAGA	AAGCTCATGA	2200
	CCAATTTCTA	GCCGACCTTG	CGAGCATTCT	ACCGAGTAAC	ACCACACCGC	2250
	TCATTGTGAG	TGATGCTGGC	TTTAAAGTGC	CATGGTATAA	ATCCGTTGAG	2300
	AAGCTGGGTT	GGTACTGGTT	AAGTCGAGTA	AGAGGAAAAG	TACAATATGC	2350
	AGACCTAGGA	GCGGAAAAC	GGAAACCTAT	CAGCAACTTA	CATGATATGT	2400
50	CATCTAGTCA	CTCAAAGACT	TTAGGCTATA	AGAGGCTGAC	TAAAAGCAAT	2450
	CCAATCTCAT	GCCAAATTCT	ATTGTATAAA	TCTCGCTCTA	AAGGCCGAAA	2500
	AAATCAGCGC	TCGACACGGA	CTCATTGTCA	CCACCCGTCA	CCTAAATCT	2550
	ACTCAGCGTC	GGCAAAGGAG	CCATGGGTTT	TAGCAACTAA	CTTACCTGTT	2600
	GAAATTCGAA	CACCCAAACA	ACTTGTTAAT	ATCTATTCTGA	AGCGAATGCA	2650
55	GATTGAAGAA	ACCTTCCGAG	ACTTGAAAAG	TCCTGCCTAC	GGACTAGGCC	2700
	TACGCCATAG	CCGAACGAGC	AGCTCAGAGC	GTTTTGATAT	CATGCTGCTA	2750
	ATCGCCCTGA	TGCTTCAACT	AACATGTTGG	CTTGCGGGCG	TTCATGCTCA	2800
	GAAACAAGGT	TGGGACAAGC	ACTTCCAGGC	TAACACAGTC	AGAAATCGAA	2850
	ACGTACTCTC	AACAGTTCGC	TTAGGCATGG	AAGTTTTGCG	GCATTCTGGC	2900
60	TACACAATAA	CAAGGGAAGA	CTTACTCGTG	GCTGCAACCC	TACTAGCTCA	2950
	AAATTTATT	ACACATGGTT	ACGCTTGGG	GAAATTATGA	TAATGATCCA	3000
	GATCACTTCT	GGCTAATAAA	AGATCAGAGC	TCTAGAGATC	TGTGTGTTGG	3050

	TTTTTTGTGG	ATCTGCTGTG	CCTTCTAGTT	GCCAGCCATC	TGTTGTTTGC	3100
	CCCTCCCCCG	TGCCTTCCTT	GACCCTGGAA	GGTGCCACTC	CCACTGTCCT	3150
	TTCTTAATAA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT	AGGTGTCATT	3200
	CTATTCTGGG	GGGTGGGGTG	GGGCAGCACA	GCAAGGGGGA	GGATTGGGAA	3250
5	GACAATAGCA	GGCATGCTGG	GGATGCGGTG	GGCTCTATGG	GTACCTCTCT	3300
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCGGTAC	CTCTCTCTCT	3350
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCAGG	TGCTGAAGAA	3400
	TTGACCCGGT	GACCAAAGGT	GCCTTTTATC	ATCACTTTAA	AAATAAAAAA	3450
	CAATTACTCA	GTGCCTGTTA	TAAGCAGCAA	TTAATTATGA	TTGATGCCTA	3500
10	CATCACAACA	AAAAC TGATT	TAACAAATGG	TTGGTCTGCC	TTAGAAAGTA	3550
	TATTTGAACA	TTATCTTGAT	TATATTATTG	ATAATAATAA	AAACCTTATC	3600
	CCTATCCAAG	AAGTGATGCC	TATCATTGGT	TGGAATGAAC	TTGAAAAAAA	3650
	TTAGCCTTGA	ATACATTACT	GGTAAGGTAA	ACGCCATTGT	CAGCAAATTG	3700
	ATCCAAGAGA	ACCAACTTAA	AGCTTTCCTG	ACGGAATGTT	AATTCTCGTT	3750
15	GACCCTGAGC	ACTGATGAAT	CCCCTAATGA	TTTTGGTAAA	AATCATTAAG	3800
	TTAAGGTGGA	TACACATCTT	GTCATATGAT	CCCGGTAATG	TGAGTTAGCT	3850
	CACTCATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	3900
	TGTGTGGAAT	TGTGAGCGGA	TAACAATTC	ACACAGGAAA	CAGCTATGAC	3950
	CATGATTACG	CCAAGCGCGC	AATTAACCC	CACTAAAGGG	AACAAAAGCT	4000
20	GGAGCTCCAC	CGCGGTGGCG	GCCGCTCTAG	AACTAGTGG	TCCCCCGGGG	4050
	AGGTCAGAA	GGTTTCTTTA	CTGTTTGTC	ATTCTATTAT	TTCAATACAG	4100
	AACAATAGCT	TCTATAACTG	AAATATATTT	GCTATTGTAT	ATTATGATTG	4150
	TCCCTCGAAC	CATGAACACT	CCTCCAGCTG	AATTTACAAA	TTCCTCTGTC	4200
	ATCTGCCAGG	CCATTAAGTT	ATTCATGGAA	GATCTTTGAG	GAACACTGCA	4250
25	AGTTCATATC	ATAAACACAT	TTGAAATGGA	GTATTGTTTT	GCATTGTATG	4300
	GAGCTATGTT	TTGCTGTATC	CTCAGAAAAA	AAGTTTGTTA	TAAAGCATTG	4350
	ACACCCATAA	AAAGATAGAT	TTAAATATTC	CAGCTATAGG	AAAGAAAGTG	4400
	CGTCTGCTCT	TCACTCTAGT	CTCAGTTGGC	TCCTTCACAT	GCATGCTTCT	4450
	TTATTTCTCC	TATTTTGTC	AGAAAATAAT	AGGTCACGTC	TTGTTCTCAC	4500
30	TTATGTCCTG	CCTAGCATGG	CTCAGATGCA	CGTTGTAGAT	ACAAGAAGGA	4550
	TCAAATGAAA	CAGACTTCTG	GTCTGTTACT	ACAACCATAG	TAATAAGCAC	4600
	ACTAACTAAT	AATTGCTAAT	TATGTTTTC	ATCTCTAAGG	TTCCACATT	4650
	TTTCTGTTTT	CTTAAAGATC	CCATTATCTG	GTGTAACTG	AAGCTCAATG	4700
	GAACATGAGC	AATATTTC	AGTCTTCTCT	CCCATCCAAC	AGTCCTGATG	4750
35	GATTAGCAGA	ACAGGCAGAA	AACACATTGT	TACCCAGAAT	TAAAACTAA	4800
	TATTTGCTCT	CCATTCAATC	CAAAATGGAC	CTATTGAAAC	TAAAATCTAA	4850
	CCCAATCCCA	TTAAATGATT	TCTATGGCGT	CAAAGGTCAA	ACTTCTGAAG	4900
	GGAACCTGTG	GGTGGGTCAC	AATTCAGGCT	ATATATTCCC	CAGGGCTCAG	4950
40	CGGATCCATG	GGCTCCATCG	GCGCAGCAAG	CATGGAATTT	TGTTTTGATG	5000
	TATTCAAGGA	GCTCAAAGTC	CACCATGCCA	ATGAGAACAT	CTTCTACTGC	5050
	CCCATTGCCA	TCATGTCAGC	TCTAGCCATG	GTATACCTGG	GTGCAAAAGA	5100
	CAGCACCAGG	ACACAGATAA	ATAAGGTTGT	TCGCTTTGAT	AACTTCCAG	5150
	GATTCGGAGA	CAGTATTGAA	GCTCAGTGTG	GCACATCTGT	AAACGTTTAC	5200
45	TCTTCACTTA	GAGACATCCT	CAACCAAATC	ACCAAACCAA	ATGATGTTTA	5250
	TTCGTTTCAGC	CTTGCCAGTA	GACTTTATGC	TGAAGAGAGA	TACCCAATCC	5300
	TGCCAGAATA	CTTGCAAGTGT	GTGAAGGAAC	TGTATAGAGG	AGGCTTGGAA	5350
	CCTATCAACT	TTCAAACAGC	TGCAGATCAA	GCCAGAGAGC	TCATCAATTC	5400
	CTGGGTAGAA	AGTCAGACAA	ATGGAATTAT	CAGAAATGTC	CTTCAGCCAA	5450
50	GCTCCGTGGA	TTCTCAAAC	GCAATGGTTC	TGGTTAATGC	CATTGTCTTC	5500
	AAAGGACTGT	GGGAGAAAAC	ATTTAAGGAT	GAAGACACAC	AAGCAATGCC	5550
	TTTCAGAGTG	ACTGAGCAAG	AAAGCAAACC	TGTGCAGATG	ATGTACCAGA	5600
	TTGGTTTATT	TAGAGTGGCA	TCAATGGCTT	CTGAGAAAAT	GAAGATCCTG	5650
	GAGCTTCCAT	TTGCCAGTGG	GACAATGAGC	ATGTTGGTGC	TGTTGCCTGA	5700
	TGAAGTCTCA	GGCCTTGAGC	AGCTTGAGAG	TATAATCAAC	TTTGAAAAAC	5750
55	TGACTGAATG	GACCAGTTCT	AATGTTATGG	AAGAGAGGAA	GATCAAAGTG	5800
	TACTTACCTC	GCATGAAGAT	GGAGGAAAAA	TACAACCTCA	CATCTGTCTT	5850
	AATGGCTATG	GGCATTACTG	ACGTGTTTAG	CTCTTCAGCC	AATCTGTCTG	5900
	GCATCTCCTC	AGCAGAGAGC	CTGAAGATAT	CTCAAGCTGT	CCATGCAGCA	5950
60	CATGCAGAAA	TCAATGAAGC	AGGCAGAGAG	GTGGTAGGGT	CAGCAGAGGC	6000
	TGGAGTGGAT	GCTGCAAGCG	TCTCTGAAGA	ATTTAGGGCT	GACCATCCAT	6050
	TCCTCTTCTG	TATCAAGCAC	ATCGCAACCA	ACGCCGTTCT	CTTCTTTGGC	6100

AGATGTGTTT CCCCTCCGCG GCCAGCAGAT GACGCACCAG CAGATGACGC 6150  
ACCAGCAGAT GACGCACCAG CAGATGACGC ACCAGCAGAT GACGCACCAG 6200  
CAGATGACGC AACAACATGT ATCCTGAAAG GCTCTGTGG CTGGATCGGC 6250  
CTGCTGGATG ACGATGACAA AAAATACAAA AAAGCACTGA AAAAAGTGGC 6300  
5 AAAACTGCTG TAATGAGGGC GCCTGGATCC AGATCACTTC TGGCTAATAA 6350  
AAGATCAGAG CTCTAGAGAT CTGTGTGTTG GTTTTTTGTG GATCTGCTGT 6400  
GCCTTCTAGT TGCCAGCCAT CTGTTGTTTG CCCCTCCCCC GTGCCTTCCT 6450  
TGACCCTGGA AGGTGCCACT CCCACTGTCC TTTCCTAATA AAATGAGGAA 6500  
ATTGCATCGC ATTGTCTGAG TAGGTGTCAT TCTATTCTGG GGGGTGGGGT 6550  
10 GGGGCAGCAC AGCAAGGGGG AGGATTGGGA AGACAATAGC AGGCATGCTG 6600  
GGGATGCGGT GGGCTCTATG GGTACCTCTC TCTCTCTCTC TCTCTCTCTC 6650  
TCTCTCTCTC TCTCTCGGTA CCTCTCTCGA GGGGGGGCCC GGTACCCAAT 6700  
TCGCCCTATA GTGAGTCGTA TTACGCGCGC TCACTGGCCG TCGTTTACA 6750  
ACGTCGTGAC TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG 6800  
15 CACATCCCCC TTTCGCCAGC TGGCGTAATA GCGAAGAGGC CCGCACCAGT 6850  
CGCCCTTCCC AACAGTTGCG CAGCCTGAAT GGCGAATGGA AATTGTAAGC 6900  
GTTAATATTT TGTAAATTT CGCGTTAAAT TTTTGTAAA TCAGCTCATT 6950  
TTTTAACCAA TAGGCCGAAA TCGGCAAAAT CCCTTATAAA TCAAAAGAAT 7000  
AGACCGAGAT AGGGTTGAGT GTTGTTCAG TTTGGAACAA GAGTCCACTA 7050  
20 TTAAAGAACG TGGACTCCAA CGTCAAAGGG CGAAAAACCG TCTATCAGGG 7100  
CGATGGCCCA CTACTCCGGG ATCATATGAC AAGATGTGTA TCCACCTTAA 7150  
CTTAATGATT TTTACCAAAA TCATTAGGGG ATTCATCAGT GCTCAGGGTC 7200  
AACGAGAATT AACATTCCGT CAGGAAAGCT TATGATGATG ATGTGCTTAA 7250  
AAACTTACTC AATGGCTGGT TATGCATATC GCAATACATG CGAAAAACCT 7300  
25 AAAAGAGCTT GCCGATAAAA AAGGCCAATT TATTGCTATT TACCGCGGCT 7350  
TTTTATTGAG CTTGAAAGAT AAATAAAATA GATAGGTTTT ATTTGAAGCT 7400  
AAATCTTCTT TATCGTAAAA AATGCCCTCT TGGGTTATCA AGAGGGTCAT 7450  
TATATTTGCG GGAATAACAT CATTTGGTGA CGAAATAACT AAGCACTTGT 7500  
CTCCTGTTTA CTCCCCTGAG CTTGAGGGGT TAACATGAAG GTCATCGATA 7550  
30 GCAGGATAAT AATACAGTAA AACGCTAAAC CAATAATCCA AATCCAGCCA 7600  
TCCCAAATTG GTAGTGAATG ATTATAAATA ACAGCAAACA GTAATGGGCC 7650  
AATAACACCG GTTGCATTGG TAAGGCTCAC CAATAATCCC TGTAAGGCAC 7700  
CTTGCTGATG ACTCTTTGTT TGGATAGACA TCACTCCCTG TAATGCAGGT 7750  
AAAGCGATCC CACCACCAGC CAATAAAATT AAAACAGGGA AAATAACCA 7800  
35 ACCTTCAGAT ATAAACGCTA AAAAGGCAAA TGCACTACTA TCTGCAATAA 7850  
ATCCGAGCAG TACTGCCGTT TTTTCGCCCC ATTTAGTGGC TATTCTTCCT 7900  
GCCACAAAGG CTTGGAATAC TGAGTGTAAG AGACCAAGAC CCGCTAATGA 7950  
AAAGCCAACC ATCATGCTAT TCCATCCAAA ACGATTTTCG GTAAATAGCA 8000  
CCACACCCGT TCGGGGAATT TGGCCTATCA ATTGCGCTGA AAAATAAATA 8050  
40 ATCAACAAAA TGGCATCGTT TTAAATAAAG TGATGTATAC CGAATTCAGC 8100  
TTTTGTTCCT TTTAGTGAGG GTTAATTGCG CGCTTGGCGT AATCATGGTC 8150  
ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT CCACACAACA 8200  
TACGAGCCGG AAGCATAAAG TGTAAGCCT GGGGTGCCTA ATGAGTGAGC 8250  
TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA 8300  
45 CCTGTGTCG CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG 8350  
GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA CTGCTGCGC 8400  
TCGGTCGTTT GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT 8450  
ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA 8500  
AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT 8550  
50 TTCCATAGGC TCCGCCCCC TGACGAGCAT CACAAAATC GACGCTCAAG 8600  
TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC 8650  
CTGGAAGCTC CCTCGTGC GC TCTCCTGTTT CGACCCTGCC GCTTACCGGA 8700  
TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC 8750  
ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT 8800  
55 GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC 8850  
TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC 8900  
AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG 8950  
AGTTCTTGAA GTGGTGGCCT AACTACGGCT AACTAGAAAG GACAGTATT 9000  
GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG 9050  
60 CTCTTGATCC GGCAACAAA CCACCGCTGG TAGCGGTGGT TTTTGTGTT 9100  
GCAAGCAGCA GATTACGCG AGAAAAAAG GATCTCAAGA AGATCCTTGT 9150



ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAAC CACGTTAAGG 9200  
GATTTTGGTC ATGAGATTAT CAAAAGGAT CTTACCTAG ATCCTTTTAA 9250  
ATTAAAAATG AAGTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG 9300  
TCTGACAGTT ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG 9350  
5 TCTATTTTCGT TCATCCATAG TTGCCTGACT CCCCCTCGTG TAGATAACTA 9400  
CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA 9450  
GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG 9500  
AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT 9550  
CTATTAAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT 9600  
10 TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC 9650  
GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA 9700  
CATGATCCCC CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCTCTCCG 9750  
ATCGTTGTCA GAAGTAAGTT GGCCGAGTG TTATCACTCA TGGTTATGGC 9800  
AGCACTGCAT AATTCTCTTA CTGTCTGACC ATCCGTAAGA TGCTTTTCTG 9850  
15 TGA CTGGTGA GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA 9900  
CCGAGTTGCT CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG 9950  
CAGAACTTTA AAAGTGCTCA TCATTGGAAA ACGTCTCTCG GGGCGAAAAC 10000  
TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT 10050  
GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG 10100  
20 AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC 10150  
GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT 10200  
TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA 10250  
AAATAAACAA ATAGGGGTTC CGCGCACATT TCCCCGAAA GTGCCAC 10297

25

SEQ ID NO:46 (pTnMod(Oval/ENT tag/P146/PA) - QUAIL)

CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50  
CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTTCGC 100  
30 TTTCTTCCCT TCCTTTCTCG CCACGTTTCG CGGCATCAGA TTGGCTATTG 150  
GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200  
CATGTCCAAC ATTACCGCCA TGTTGACATT GATTATTGAC TAGTTATTAA 250  
TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300  
CGTTACATAA CTTACGGTAA ATGGCCCCGCC TGGCTGACCG CCCAACGACC 350  
35 CCGGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400  
GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAAC TGCCCA 450  
CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC CCTATTGACG 500  
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA 550  
TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600  
40 CATGGTGATG CGGTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG 650  
ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700  
TTTTGGCACC AAAATCAACG GGAATTTCCA AAATGTCGTA ACAACTCCGC 750  
CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 800  
GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850  
45 TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900  
GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA 950  
CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000  
CTGTTTTTGG CTTGGGGCCT ATACACCCCC GCTTCCTTAT GCTATAGGTG 1050  
ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCATT ATTGACCACT 1100  
50 CCCCTATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150  
GCCACAATA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTGAGAGAC 1200  
TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATTATTTACA 1250  
AATTACATA TACAACAACG CCGTCCCCCG TGCCCGCAGT TTTTATTAAA 1300  
CATAGCGTGG GATCTCCACG CGAATCTCGG GTACGTGTTC CGGACATGGG 1350  
55 CTCTTCTCCG GTAGCGGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400  
CTCCAGCGGC TCATGGTCGC TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450  
CCAGACTTAG GCACAGCACA ATGCCACCA CCACCAGTGT GCCGCACAAG 1500  
GCCGTGGCGG TAGGGTATGT GTCTGAAAAT GAGCGTGGAG ATTGGGCTCG 1550  
CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600  
60 GCAGCTGAGT TGTGTATTTC TGATAAGAGT CAGAGGTAAC TCCCGTTGCG 1650  
GTGCTGTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700

CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCTTTT 1750  
CCATGGGTCT TTTCTGCAGT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800  
TTACATGATT CTCTTTACCA ATTCTGCCCC GAATTACACT TAAAACGACT 1850  
CAACAGCTTA ACGTTGGCTT GCCACGCATT ACTTGACTGT AAAACTCTCA 1900  
5 CTCTTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCGAG AACAAAACAT 1950  
AACATCAAAC GAATCGACCG ATTGTTAGGT AATCGTCACC TCCACAAAGA 2000  
GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050  
GATGCCCATT GTACTTGTG ACTGGTCTGA TATTCGTGAG CAAAAACGAC 2100  
TTATGGTATT GCGAGCTTCA GTCGCACTAC ACGGTCGTTT TGTACTCTT 2150  
10 TATGAGAAAG CGTTCCCGCT TTCAGAGCAA TGTTCAAAGA AAGCTCATGA 2200  
CCAATTCTA GCCGACCTTG CGAGCATTCT ACCGAGTAAC ACCACACCGC 2250  
TCATTGTCAG TGATGCTGGC TTTAAAGTGC CATGGTATAA ATCCGTTGAG 2300  
AAGCTGGGT GGTACTGGTT AAGTCGAGTA AGAGGAAAAG TACAATATGC 2350  
AGACCTAGGA GCGGAAAAC GGAACCTAT CAGCAACTTA CATGATATGT 2400  
15 CATCTAGTCA CTCAAAGACT TTAGGCTATA AGAGGCTGAC TAAAAGCAAT 2450  
CCAATCTCAT GCCAAATCT ATTGTATAAA TCTCGCTCTA AAGGCCGAAA 2500  
AAATCAGCGC TCGACACGGA CTCATTGTCA CCACCCGTCA CCTAAAATCT 2550  
ACTCAGCGTC GGCAAAGGAG CCATGGGTTC TAGCAACTAA CTTACCTGTT 2600  
GAAATTCGAA CACCCAAACA ACTTGTTAAT ATCTATTCTGA AGCGAATGCA 2650  
20 GATTGAAGAA ACCTTCCGAG ACTTGAAAAG TCCTGCCTAC GGACTAGGCC 2700  
TACGCCATAG CCGAACGAGC AGCTCAGAGC GTTTTGATAT CATGCTGCTA 2750  
ATCGCCCTGA TGCTTCAACT AACATGTTGG CTTGCGGGCG TTCATGCTCA 2800  
GAAACAAGGT TGGGACAAGC ACTTCCAGGC TAACACAGTC AGAAATCGAA 2850  
ACGTACTCTC AACAGTTCGC TTAGGCATGG AAGTTTTGCG GCATTCTGGC 2900  
25 TACACAATAA CAAGGGAAGA CTTACTCGTG GCTGCAACCC TACTAGCTCA 2950  
AAATTTATTC ACACATGGTT ACGCTTTGGG GAAATTATGA TAATGATCCA 3000  
GATCACTTCT GGCTAATAAA AGATCAGAGC TCTAGAGATC TGTGTGTTGG 3050  
TTTTTTGTGG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC TGTGTTTGC 3100  
CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC CCACTGTCCT 3150  
30 TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT 3200  
CTATTCTGGG GGGTGGGGTG GGGCAGCACA GCAAGGGGGA GGATTGGGAA 3250  
GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCTCTCT 3300  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCGGTAC CTCTCTCTCT 3350  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CGGTACCAGG TGCTGAAGAA 3400  
35 TTGACCCGGT GACCAAAGGT GCCTTTTATC ATCACTTTAA AAATAAAAAA 3450  
CAATTACTCA GTGCCTGTTA TAAGCAGCAA TTAATTATGA TTGATGCCTA 3500  
CATCACAACA AAAACTGATT TAACAAATGG TTGGTCTGCC TTAGAAAGTA 3550  
TATTTGAACA TTATCTTGAT TATATTATTG ATAATAATAA AAACCTTATC 3600  
CCTATCCAAG AAGTGATGCC TATCATTGGT TGGAATGAAC TTGAAAAAAA 3650  
40 TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATTG 3700  
ATCCAAGAGA ACCAACTTAA AGCTTTCCTG ACGGAATGTT AATTCTCGTT 3750  
GACCCTGAGC ACTGATGAAT CCCCTAATGA TTTTGGTAAA AATCATTAAG 3800  
TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850  
CACTCATTAG GCACCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT 3900  
45 TGTGTGGAAT TGTGAGCGGA TAACAATTC ACACAGGAAA CAGCTATGAC 3950  
CATGATTACG CCAAGCGCGC AATTAACCTT CACTAAAGGG AACAAAAGCT 4000  
GGAGCTCCAC CGCGGTGGCG GCCGCTCTAG AACTAGTGA TCCCCGGGG 4050  
AGGTCAGAAT GGTTTCTTTA CTGTTTGTCA ATTCTATTAT TTCAATACAG 4100  
AACAAAAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG 4150  
50 TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTCACAA TTCTCTGTC 4200  
ATCTGCCAGG CTGGAAGATC ATGGAAGATC TCTGAGGAAC ATTGCAAGTT 4250  
CATACCATAA ACTCATTGGA AATTGAGTAT TATTTTGCTT TGAATGGAGC 4300  
TATGTTTTGC AGTTCCCTCA GAAGAAAAGC TTGTTATAAA GCGTCTACAC 4350  
CCATCAAAAG ATATATTTAA ATATTCCAAC TACAGAAAGA TTTTGTCTGC 4400  
55 TCTTCACTCT GATCTCAGTT GGTTTCTTCA CGTACATGCT TCTTTATTTG 4450  
CCTATTTTGT CAAGAAAATA ATAGGTCAAG TCCTGTTCTC ACTTATCTCC 4500  
TGCCTAGCAT GGCTTAGATG CACGTTGTAC ATTCAAGAAG GATCAAATGA 4550  
AACAGACTTC TGGTCTGTTA CAACAACCAT AGTAATAAAC AGACTAACTA 4600  
ATAATTGCTA ATTATGTTTT CCATCTCTAA GGTTCCCAACA TTTTCTGTT 4650  
60 TTAAGATCCC ATTATCTGGT TGTAAGTGA GCTCAATGGA ACATGAACAG 4700  
TATTTCTCAG TCTTTTCTCC AGCAATCCTG ACGGATTAGA AGAACTGGCA 4750

5 GAAAACACTT TGTTACCCAG AATTAAAAAC TAATATTTGC TCTCCCTTCA 4800  
ATCCAAAATG GACCTATTGA AACTAAAATC TGACCCAATC CCATTAAATT 4850  
ATTTCTATGG CGTCAAAGGT CAAACTTTTG AAGGGAACCT GTGGGTGGGT 4900  
CCCAATTCAG GCTATATATT CCCCAGGGCT CAGCCAGTGG ATCCATGGGC 4950  
TCCATCGGTG CAGCAAGCAT GGAATTTTGT TTTGATGTAT TCAAGGAGCT 5000  
CAAAGTCCAC CATGCCAATG ACAACATGCT CTAATCCCCC TTTGCCATCT 5050  
TGTCAACTCT GGCCATGGTC TTCCTAGGTG CAAAAGACAG CACCAGGACC 5100  
CAGATAAATA AGGTTGTTCA CTTTGATAAA CTTCCAGGAT TCGGAGACAG 5150  
TATTGAAGCT CAGTGTGGCA CATCTGTAAA TGTTCACTCT TCACTTAGAG 5200  
10 ACATACTCAA CCAAATCACC AAACAAAATG ATGCTTATTC GTTCAGCCTT 5250  
GCCAGTAGAC TTTATGCTCA AGAGACATAC ACAGTCGTGC CGGAATACTT 5300  
GCAATGTGTG AAGGAACTGT ATAGAGGAGG CTTAGAATCC GTCAACTTTC 5350  
AAACAGCTGC AGATCAAGCC AGAGGCCTCA TCAATGCCTG GGTAAGAAAGT 5400  
CAGACAAACG GAATTATCAG AAACATCCTT CAGCCAAGCT CCGTGGATTC 5450  
15 TCAAAC TGCA ATGGTCCTGG TTAATGCCAT TGCCTTCAAG GGACTGTGGG 5500  
AGAAAGCATT TAAGGCTGAA GACACGCAAA CAATACCTTT CAGAGTGAAT 5550  
GAGCAAGAAA GCAAACCTGT GCAGATGATG TACCAGATTG GTTCATTTAA 5600  
AGTGGCATCA ATGGCTTCTG AGAAAATGAA GATCCTGGAG CTTCCATTTG 5650  
CCAGTGGAAC AATGAGCATG TTGGTGCTGT TGCCTGATGA TGTCTCAGGC 5700  
20 CTTGAGCAGC TTGAGAGTAT AATCAGCTTT GAAAACTGA CTGAATGGAC 5750  
CAGTTCTAGT ATTATGGAAG AGAGGAAGGT CAAAGTGTAC TTACCTCGCA 5800  
TGAAGATGGA GGAGAAATAC AACCTCACAT CTCTCTTAAT GGCTATGGGA 5850  
ATTACTGACC TGTTCAAGTC TTCAGCCAAT CTGTCTGGCA TCTCCTCAGT 5900  
AGGGAGCCTG AAGATATCTC AAGCTGTCCA TGCAGCACAT GCAGAAATCA 5950  
25 ATGAAGCGGG CAGAGATGTG GTAGGCTCAG CAGAGGCTGG AGTGGATGCT 6000  
ACTGAAGAAT TTAGGGCTGA CCATCCATTC CTCTTCTGTG TCAAGCACAT 6050  
CGAAACCAAC GCCATTCTCC TCTTTGGCAG ATGTGTTTCT CCGCGGCCAG 6100  
CAGATGACGC ACCAGCAGAT GACGCACCAG CAGATGACGC ACCAGCAGAT 6150  
GACGCACCAG CAGATGACGC ACCAGCAGAT GACGCAACAA CATGTATCCT 6200  
30 GAAAGGCTCT TGTGGCTGGA TCGGCCTGCT GGATGACGAT GACAAAAAAT 6250  
ACAAAAAAGC ACTGAAAAAA CTGGCAAAAC TGCTGTAATG AGGGCGCCTG 6300  
GATCCAGATC ACTTCTGGCT AATAAAAGAT CAGAGCTCTA GAGATCTGTG 6350  
TGTTGGTTTT TTGTGGATCT GCTGTGCCTT CTAGTTGCCA GCCATCTGTT 6400  
GTTTGCCCTT CCCCCGTGCC TTCCTTGACC CTGGAAGGTG CCACTCCCAC 6450  
35 TGTCCTTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT CTGAGTAGGT 6500  
GTCATTCTAT TCTGGGGGGT GGGGTGGGGC AGCACAGCAA GGGGGAGGAT 6550  
TGGGAAGACA ATAGCAGGCA TGCTGGGGAT GCGGTGGGCT CTATGGGTAC 6600  
CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CCGTACCTCT 6650  
CTCGAGGGGG GGCCCGGTAC CCAATTGCGC CTATAGTGAG TCGTATTACG 6700  
40 CGCGCTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA AAACCCTGGC 6750  
GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTCTCG CCAGCTGGCG 6800  
TAATAGCGAA GAGGCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC 6850  
TGAATGGCGA ATGGAAATTG TAAGCGTTAA TATTTTGTTA AAATTCGCGT 6900  
TAAATTTTTG TTAAATCAGC TCATTTTTTA ACCAATAGGC CGAAATCGGC 6950  
45 AAAATCCCTT ATAAATCAAA AGAATAGACC GAGATAGGGT TGAGTGTGTG 7000  
TCCAGTTTGG AACAAAGATC CACTATTAAA GAACGTGGAC TCCAACGTCA 7050  
AAGGGCGAAA AACCCTCTAT CAGGGCGATG GCCCACTACT CCGGGATCAT 7100  
ATGACAAGAT GTGTATCCAC CTTAACTTAA TGATTTTTAC CAAAATCATT 7150  
AGGGGATTCA TCAGTGCTCA GGGTCAACGA GAATTAACAT TCCGTCAGGA 7200  
50 AAGCTTATGA TGATGATGTG CTTAAAAACT TACTCAATGG CTGGTTATGC 7250  
ATATCGCAAT ACATGCGAAA AACCTAAAAG AGCTTGCCGA TAAAAAAGGC 7300  
CAATTTATTG CTATTTACCG CGGCTTTTTA TTGAGCTTGA AAGATAAATA 7350  
AAATAGATAG GTTTTATTTG AAGCTAAATC TTCTTTATCG TAAAAAATGC 7400  
CCTCTGGGT TATCAAGAGG GTCATTATAT TTCGCGGAAT AACATCATT 7450  
55 GGTGACGAAA TAACTAAGCA CTTGTCTCCT GTTTACTCCC CTGAGCTTGA 7500  
GGGGTTAACA TGAAGGTCAT CGATAGCAGG ATAATAATAC AGTAAACGC 7550  
TAAACCAATA ATCCAAATCC AGCCATCCCA AATTGGTAGT GAATGATTAT 7600  
AAATAACAGC AAACAGTAAT GGGCCAATAA CACCGGTTGC ATTGGTAAGG 7650  
CTCACCATA ATCCCTGTAA AGCACCTTGC TGATGACTCT TTGTTTGGAT 7700  
60 AGACATCACT CCCTGTAAATG CAGGTAAAGC GATCCCACCA CCAGCCAATA 7750  
AAATTAAAC AGGGAAAAC AACCACCTT CAGATATAAA CGCTAAAAAG 7800



5 GCAAATGCAC TACTATCTGC AATAAATCCG AGCAGTACTG CCGTTTTTTC 7850  
GCCCCATTTA GTGGCTATTC TTCCTGCCAC AAAGGCTTGG AATACTGAGT 7900  
GTAAAAGACC AAGACCCGCT AATGAAAAGC CAACCATCAT GCTATTCAT 7950  
CCAAAACGAT TTTCGGTAAA TAGCACCCAC ACCGTTGCGG GAATTTGGCC 8000  
TATCAATTGC GCTGAAAAAT AAATAATCAA CAAAATGGCA TCGTTTTAAA 8050  
TAAAGTGATG TATACCGAAT TCAGCTTTTG TTCCCTTTAG TGAGGGTTAA 8100  
TTGCGCGCTT GCGGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT 8150  
TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTA 8200  
AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT 8250  
10 CACTGCCCCG TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA 8300  
ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC 8350  
TTCCTCGCTC ACTGACTCGC TCGCTCGGT CGTTCGGCTG CCGCGAGCGG 8400  
TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT 8450  
AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG 8500  
15 TAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG 8550  
AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA 8600  
CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TCGCTCTCC 8650  
TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG 8700  
GAAGCGTGGC GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG 8750  
20 TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC 8800  
CGACCGCTGC GCCTTATCCG GTAACATATCG TCTTGAGTCC AACCCGGTAA 8850  
GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA 8900  
GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA 8950  
CGGCTACACT AGAAGGACAG TATTGGTAT CTGCGCTCTG CTGAAGCCAG 9000  
25 TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC 9050  
GCTGGTAGCG GTGGTTTTTT TGTTCGAAG CAGCAGATTA CGCGCAGAAA 9100  
AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC 9150  
AGTGGAACGA AAATCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA 9200  
AGGATCTTCA CCTAGATCCT TTAAATTA AATGAAGTT TTAAATCAAT 9250  
30 CTAAAGTATA TATGAGTAAA CTGGTCTGA CAGTTACCAA TGCTTAATCA 9300  
GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTTCATC CATAGTTGCC 9350  
TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG 9400  
CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT 9450  
TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT 9500  
35 GCAACTTTAT CCGCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG 9550  
AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA 9600  
CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC 9650  
GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAA 9700  
AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTGAGAAGT AAGTTGGCCG 9750  
40 CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC 9800  
ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC 9850  
ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA 9900  
TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT 9950  
45 GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG 10000  
ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT 10050  
TTACTTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC 10100  
GCAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT 10150  
CCTTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG 10200  
GATACATaTT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCGCGC 10250  
50 ACATTTCCCC GAAAAGTGCC AC 10272

SEQ ID NO:47 pTnMCS (CMV-CHOVg-ent-ProInsulin-synPA)

55 1 ctgacgcgcc ctgtageggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga  
61 ccgctacact tgcagegcc ctagegccc ctcctttcgc tttcttcctt tcctttctcg  
121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa  
181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac  
241 tagttattaa tagtaatcaa ttacgggggc attagttcat agcccatata tggagttccg  
301 cgttacataa cttacggtaa atggcccgc tggctgaccg cccaacgacc cccgcccatt  
60 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca  
421 atgggtggag tatttacggg aaactgccc cttggcagta catcaagtgt atcatatgcc  
481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcotggcatt atgcccagta

541 catgacotta tgggaactttc ctacttggca gtacatctac gtattagtca tgcgtattac  
601 catggtgatg cgggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg  
661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg  
721 ggactttcca aaatgtcgta acaactccgc ccatttgacg caaatgggag gtaggcgtgt  
781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg  
841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg  
901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag  
961 actctatagg cacaccctt tggctcttat gcatgctata ctgttttttg cttggggcct  
1021 atacaccccc gcttccttat gctatagggt atggtatagc ttagcctata ggtgtgggtt  
1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac  
1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac  
1201 tgacacggac tctgtatatt tacaggatgg ggtcccattht attatttaca aattcacata  
1261 tacaacaacg cgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg  
1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtagcggcgg agcttccaca  
1381 tccgagccct ggtcccatgc ctccagcggc tcatggctcg tcggcagctc cttgtccta  
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag  
1501 gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag attgggctcg cacggtgac  
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc  
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cgggtggagg cagtgtagtc  
1681 tgagcagtac tegtgtgtgc cgcgcgcgcc accagacata atagctgaca gactaacaga  
1741 ctgttctctt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt  
1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta  
1861 acgttggtct gccacgcatt acttgactgt aaaactctca ctcttaccga acttggcctg  
1921 aacctgcaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt  
1981 aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcagctag ctttatctgt  
2041 tcgggcaata cgatgcccat tgtacttgtt gactggctctg atattcgtga gcaaaaacga  
2101 cttatggtat tgcgagcttc agtgcacta cacggctcgt ctgttactct ttatgagaaa  
2161 gcgttcccg cttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt  
2221 gcgagcattc taccgagtaa caccacaccg ctcatgttca gtgatgctgg ctttaaagtg  
2281 ccatgggtata aatccgttga gaagctgggt tggactgggt taagtcgagt aagaggaaaa  
2341 gtacaatatg cagacctagg agcggaaaac tggaaaacct tcagcaactt acatgatatg  
2401 tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca  
2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg  
2521 actcattgtc accaccctgc acctaaaac tactcagcgt cggcaaagga gccatgggtt  
2581 ctagcaacta acttacctgt tgaaattcga acacccaaac aacttgtaa tatctattcg  
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc  
2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg  
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag  
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg  
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggtgtcaacc  
2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg aggggatcgc  
3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaagggtgcct tttatcatca  
3061 ctttaaaaat aaaaaacaat tactcagtcg ctgttataag cagcaattaa ttatgattga  
3121 tgccctacatc acaacaaaaa ctgatttaac aaatgggttg tctgccttag aaagtatatt  
3181 tgaacattat cttgattata ttattgataa taataaaaaa cttatcccta tccaagaagt  
3241 gatgcctatc attggttggg atgaacttga aaaaaattag ccttgaatac attactggta  
3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg  
3361 aatgttaatt ctcgttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc  
3421 attaagttaa ggtggataca catcttgtca tatgatcccc gtaatgtgag ttagctcact  
3481 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tggaattgtg  
3541 agcggataac aatttcacac aggaaacagc tatgaccatg attacgcaa gcgcgcaatt  
3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcggcgg ctctagaact  
3661 agtggatccc cggggcatca gattggctat tggccattgc atacgttgta tccatatcat  
3721 aatatgtaca tttatattgg ctcatgtcca acattaccgc catgttgaca ttgattattg  
3781 actagttatt aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc  
3841 cgcgttacat aacttacggt aaatggccc cctggctgac cgcceaacga cccccccca  
3901 ttgacgtcaa taatgacgta tgttcccata gtaacgcaa tagggacttt ccattgacgt  
3961 caatgggtgg agtatttacg gtaacttgc cacttggcag tacatcaagt gtatcatatg  
4021 ccaagtacgc cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgccag  
4081 tacatgacct tatgggactt tctacttgg cagtacatct acgtattagt catcgctatt  
4141 accatgggtg tgcggttttg gcagtacatc aatgggcgtg gatagcgtt tgactcacgg  
4201 ggatttccaa gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa  
4261 cgggactttc caaatgtcg taacaactcc gcccattga cgcaaatggg cggtaggcgt  
4321 gtacgggtgg aggtctatat aagcagagct cgttttagtg accgtcagat cgcctggaga  
4381 cgccatccac gctgttttga cctccataga agacaccggg accgatccag cctccgcggc  
4441 cgggaacggt gcattggaac gcggttccc cgtgccaaga gtgacgtaag taccgcctat  
4501 agactctata ggcacacccc tttggctctt atgcatgcta tactgttttt ggcttggggc  
4561 ctatacaccc ccgcttcctt atgctatagg tgatgggtata gcttagccta taggtgtggg

5 4621 ttattgacca ttattgacca ctccccattt ggtgacgata ctttccatta ctaatccata  
4681 acatggctct ttgccacaac tatctctatt ggctatatgc caatactctg tccctcagag  
4741 actgacacgg actctgtatt tttacaggat ggggtcccat ttattattta caaattcaca  
4801 tataacaaca cgccgtcccc cgtgcccga gtttttatta aacatagcgt gggatctcca  
4861 cgcgaatctc gggtagctgt tccggacatg ggctctctc cggtagcggc ggagcttcca  
4921 catccgagcc ctgggtcccat gctccagcg gctcatggtc gctcggcagc tccctgctcc  
4981 taacagtggg ggccagactt aggcacagca caatgccac caccaccagt gtgcccagca  
5041 aggccgtggc ggtagggat gtgtctgaaa atgagcgtgg agattgggt cgcacggctg  
10 5101 acgcagatgg aagacttaag gcagcggcag aagaagatgc aggcagctga gttgtgtat  
5161 tctgataaga gtcagaggta actcccgttg cgggtgtgtt aacgggtggag ggcagtgtag  
5221 tctgagcagt actcgttgct gccgcgcgcg ccaccagaca taatagctga cagactaaca  
5281 gactgttctt tccatgggt cttttctgca gtcaccgtcg gatcaatggg cccatcggt  
5341 gcagcaagca tgggaattttg ttttgatgta ttcaaggagc tcaaagtcca ccatgccaat  
5401 gagaacatct tctactgcc cattgccatc atgtcagctc tagccatggg atacctgggt  
15 5461 gcaaaagaca gcaccaggac acaataaat aaggttggtc gctttgataa acttccagga  
5521 ttccggagaca gtattgaagc tcagtgtggc acatctgtaa acgttccactc ttcacttaga  
5581 gacatcctca accaaatcac caaaccaaat gatgtttatt cgttcagcct tgcagtaga  
5641 ctttatgctg aagagagata cccaatcctg ccagaatact tgcagtgtgt gaaggaaactg  
5701 tatagaggag gcttggaaacc tatcaacttt caaacagctg cagatcaagc cagagagctc  
20 5761 atcaattcct gggtagaaag tcagacaaat ggaattatca gaaatgtcct tcagccaagc  
5821 tccgtggatt ctcaaatcgc aatggttctg gttaatgcca ttgtcttcaa aggactgtgg  
5881 gagaaagcat ttaaggatga agacacacaa gcaatgcctt tcagagtac tgagcaagaa  
5941 agcaaacctg tgcagatgat gtaccagatt ggtttattta gagtggcatc aatggcttct  
6001 gagaaaatga agatcctgga gcttccattt gccagtggga caatgagcat gttggtgctg  
25 6061 ttgcctgatg aagtctcagg ccttgagcag cttgagagta taatcaactt tgaaaaactg  
6121 actgaatgga ccagttctaa tgttatggaa gagagaagat caaagtgtac ttacctcgca  
6181 tgaagatgga ggaataatc aacctcacat ctgtcttaab ggctatggc attactgacg  
6241 tgtttagctc ttcagccaat ctgtctggca tctcctcagc agagagcctg aagatattctc  
6301 aagctgtcca tgcagcacat gcagaaatca atgaagcagg cagagagggtg gtagggtcag  
30 6361 cagaggctgg agtggtatgt gcaagcgtct ctgaagaatt tagggctgac catccattcc  
6421 tcttctgtat caagcacatc gcaaccaacg ccgttctctt cttttggcag atgtgttcc  
6481 cgcggccagc agatgacgca ccagcagatg acgcaccagc agatgacgca ccagcagatg  
6541 acgcaccagc agatgacgca acaacatgta tcttgaaagg ctcttggtgc tggatcgggc  
6601 tgctggatga cgatgacaaa tttgtgaacc aacacctgtg cggctcacac ctggtggaag  
35 6661 ctctctacct agtgtgcggg gaacgaggct tcttctacac acccaagacc cgcggggagg  
6721 cagaggacct gcaggtgggg caggtggagc tgggcggggg ccctggtgca ggcagcctgc  
6781 agcccttggc cctggagggg tccctgcaga agcgtggcat tgtggaacaa tgctgtacca  
6841 gcattctgct cctctaccag ctggagaact actgcaacta gggcgccctaa agggcgaaat  
6901 atcgcggccg ctctagacca ggcgcctgga tccagatcac ttctggctaa taaaagatca  
40 6961 gagetctaga gatctgtgtg ttgggttttt gtggatctgc tgtgccttct agttgccagc  
7021 catctgttgt ttgcccctcc cccgtgcctt ccttgaccct ggaagggtgc actcccactg  
7081 tcccttccta ataaaatgag gaaattgcat cgcattgtct gagtagggtg cattctattc  
7141 tgggggggtg ggtggggcag cacagcaagg gggaggattg ggaagacaat agcaggcatg  
7201 ctggggatgc ggtgggtct atgggtacct ctctctctct ctctctctct ctcactctct  
45 7261 ctctctctcg gtacctctcc tcgagggggg gccgggtacc caattcgccc tatagtagt  
7321 cgtattacgc gcgtcactg gccgtcgttt tacaacgtcg tgactgggaa aacctggcg  
7381 ttacccaact taatgcctt gcagcacatc ccccttctgc cagctggcgt aatagcgaag  
7441 agggccgcac cgatgcctt tcccaacagt tgcgcagcct gaatggcgaa tggaaattgt  
50 7501 aagcgttaat attttgttaa aattcgcgtt aaatttttgt taaatcagct cattttttta  
7561 ccaataggcc gaaatcgcca aaatccctta taaatcaaaa gaatagaccg agatagggtt  
7621 gagtgttgtt ccagtttgga acaagagtc actattaaag aacgtggact ccaacgtcaa  
7681 agggcgaaaa accgtctatc agggcgatgg cccactactc cgggatcata tgacaagatg  
7741 tgtatccacc ttaacttaat gatttttacc aaaatcatta ggggattcat cagtgtcag  
55 7801 ggtcaacgag aattaacatt ccgtcaggaa agcttatgat gatgatgtgc ttaaaaactt  
7861 actcaatggc tggttatgca tatcgcaata catgcgaaaa acctaaaaga gcttgccgat  
7921 aaaaaaggcc aatttattgc tatttaccgc ggctttttat tgagcttgaa agataaataa  
7981 aatagatagg ttttatttga agctaaatct tctttatcgt aaaaaatgcc ctcttgggtt  
8041 atcaagaggg tcattatatt tcgcggaata acatcatttg gtgacgaaat aactaagcac  
60 8101 ttgtctcctg tttactcccc tgagcttgag ggggttaacat gaaggctcat gatagcagga  
8161 taataatata gtaaaacgct aaaccaataa tccaaatcca gccatcccaa attggtagtg  
8221 aatgattata aataacagca aacagtaatg ggccaataac accggttgca ttggtaaggc  
8281 tcaccaataa tccctgtaaa gcaccttgct gatgactott tgtttggata gacatcactc  
8341 cctgtaatgc aggtaaagcg atccccacc cagccaataa aattaaaaca gggaaaacta  
65 8401 accaaccttc agatataaac gctaaaaagg caaatgcact actatctgca ataaatccga  
8461 gcagtactgc cgtttttctg cccatttagt ggctattctt cctgccacaa aggcttgga  
8521 tactgagtgt aaaagaccaa gaccgtaat gaaaagccaa ccatcatgct attcatcatc  
8581 acgatttctg taatagcacc acaccgtgct ggattggcta tcaatgcgct gaaataataa  
8641 tcaacaaatg gcacgttaa ataagtatg tataccgac agcttttgtt ccttttagtg



5 8701 aggggttaatt gcgcgcttgg cgtaatcatg gtcatactg tttcctgtgt gaaattgtta  
 8761 tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag cctgggggtgc  
 8821 ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg  
 8881 aaacctgtcg tgccagctgc attaatgaat .cggccaaacgc gcggggagag gcggtttgcg  
 8941 tattgggcgc tcttcgctt cctcgctcac tgactcgctg cgctcggtcg ttcggtgcg  
 9001 gcgagcggtg tcagctcact caaaggcggt aatacggtta tccacagaat .caggggataa  
 9061 cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc  
 9121 gttgctggcg tttttccata ggctccgccc cctgacgag catcacaaaa atcgacgctc  
 9181 aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag  
 10 9241 ctccctcgctg cgctctcctg ttccgacctt gccgcttacc ggatacctgt ccgctttct  
 9301 cccttcggga agcgtggcgc tttctcatag ctacgctgt aggtatctca gttcgggtga  
 9361 ggtcggtcgc tccaagctgg gctgtgtgca cgaaccccc gttcagccc accgctgcgc  
 9421 cttatccggt aactatcgct ttgagtccaa cccggtatga cagacttat cgcactggc  
 9481 agcagccact ggtaacagga ttagcagagc aggtatgta ggcggtgcta cagagtctt  
 15 9541 gaagtgggtg cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct  
 9601 gaagccagtt accttcggaa aaagagtgtg tagctctga tccggcaaac aaaccaccgc  
 9661 tggtagcgtt gggttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca  
 9721 agaagatcct ttgatctttt ctacgggtgc tgacgctcag tggaacgaaa actcacgta  
 9781 agggattttg gtcagtgcgt tatcaaaaag gatcttcacc tagatcctt taaattaaaa  
 20 9841 atgaagtttt aaatcaatct aaagtatata tgagtaaact tggcttgaca gttaccaatg  
 9901 cttaatcagt gaggcacct tctcagcgat ctgtctattt cgttcacca tagttgcctg  
 9961 actccccgct gtgtagataa ctacgatacg ggagggtta ccatctggcc ccagtgcctg  
 10021 aatgataccg cgagaccac gctcaccggc tccagattta tcagcaataa accagccagc  
 10081 cggaaggggc gagcgagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa  
 25 10141 ttgttgctcg gaagctagag taagtgttc gccagttaat agtttgccga acgttggtgc  
 10201 cattgctaca ggcacgtgg tgcacgctc gtcgtttggt atggcttcat tcagctccgg  
 10261 ttcccaacga tcaaggcgag ttacatgac cccatgttg tgcaaaaaag cggtagctc  
 10321 cttcggtcct ccgatcggtg tcagaagtaa gttggcgcga gtgttatcac tcatggttat  
 10381 ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg  
 30 10441 tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt gctcttgcce  
 10501 ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg  
 10561 aaaacgttct tcggggcgaa aactctcaag gatcttacc ctgttgagat ccagttcgat  
 10621 gtaaccact cgtgcacca actgatcttc agcatctttt actttacca gcgtttctgg  
 10681 gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgca cagggaaatg  
 35 10741 ttgaatactc atactcttc tttttcaata ttattgaagc atttatcagg gttattgtct  
 10801 catgagcgga tacatatttg aatgtattta gaaaaataa caaatagggg ttccgcgcac  
 10861 atttccccga aaagtgcac

## 40 SEQ ID NO:48 (cecropin prepro)

AAT TTC TCA AGG ATA TTT  
 TTC TTC GTG TTC GCT TTG  
 GTT CTG GCT TTG TCA ACA  
 GTT TCG GCT GCG CCA GAG  
 45 CCG AAA

## SEQ ID NO:49 (cecropin prepro extended)

50 AAT TTC TCA AGG ATA TTT  
 TTC TTC GTG TTC GCT TTG  
 GTT CTG GCT TTG TCA ACA  
 GTT TCG GCT GCG CCA GAG  
 CCG AAA TGG AAA GTC TTC  
 55 AAG

## SEQ ID NO:50 (cecropin pro)

GCG CCA GAG CCG AAA

60

SEQ ID NO:51 (cecropin pro extended)  
 GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

65

SEQ ID NO:52 (a Kozak sequence)  
 ACCATGT

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**